



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 823 479 A3

(12) **EUROPEAN PATENT APPLICATION**

(88) Date of publication A3:
29.03.2000 Bulletin 2000/13

(51) Int. Cl.7: C12N 15/54, C12N 9/12,
C12N 15/10, C12Q 1/68

(43) Date of publication A2:
11.02.1998 Bulletin 1998/07

(21) Application number: 97113182.6

(22) Date of filing: 31.07.1997

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE

(30) Priority: 06.08.1996 US 23376 P

(71) Applicant:
F. HOFFMANN-LA ROCHE AG
4070 Basel (CH)

(72) Inventors:
• Geifland, David Harrow
Oakland, CA 94611 (US)

• Kalman, Lisa Vivian
San Francisco, CA 94131 (US)
• Reichert, Fred Lawrence
Oakland, CA 94610 (US)

(74) Representative:
Wächter, Dieter Ernst, Dr. et al
F.Hoffmann-La Roche AG
Patent Department (PLP),
124 Grenzacherstrasse
4070 Basel (CH)

(54) **Modified thermostable DNA polymerase**

(57) The invention provides thermostable DNA polymerase enzymes that comprises the amino acid sequence SerGlnIleXaaLeuArgXaa (SEQ ID NO: 1), wherein "Xaa" at position 4 of this sequence is any amino acid residue but not a glutamic acid residue (Glu), preferably a glycine residue and "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile). The thermostable DNA polymerases of the invention have enhanced efficiency for incorporating unconventional nucleotides, such as ribonucleotides, into DNA products and are advantageous in many *in vitro* synthesis applications. Such enzymes are particularly useful for use in nucleic acid sequencing protocols and provide novel means for DNA sequence analysis with cost and efficiency advantages. Also claimed are nucleic acids encoding said polymerases, vectors and host cells comprising such a nucleic acid, as well as compositions for use in a DNA sequencing reaction, kits and methods for sequencing including such polymerases.



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 97 11 3182

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X, D	EP 0 655 506 A (HARVARD COLLEGE) 31 May 1995 (1995-05-31) * examples 6-20 * ---	2-22	C12N15/54 C12N9/12 C12N15/10 C1201/68
X	TABOR S ET AL: "A SINGLE RESIDUE IN DNA POLYMERASES I OF THE ESCHERICHIA COLI DNA POLYMERASE I FAMILY IS CRITICAL FOR DISTINGUISHING BETWEEN DEOXY- AND DIDEOXYRIBONUCLEOTIDES" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, US, NATIONAL ACADEMY OF SCIENCE, WASHINGTON, vol. 92, 1 July 1995 (1995-07-01), pages 6339-6343, XP000606203 ISSN: 0027-8424 *the abstract* ---	2-22	
X	WO 96 10640 A (LIFE TECHNOLOGIES INC) 11 April 1996 (1996-04-11) * page 29, line 25 - line 30; example 15 * ---	2,3,7-15	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)
			C12N
The present search report has been drawn up for all claims			
Piece of search	Date of completion of the search	Examiner	
THE HAGUE	7 February 2000	Cupido, M	
CATEGORY OF CITED DOCUMENTS			
X: particularly relevant if taken alone Y: particularly relevant if combined with another document of the same category A: technological background O: non-patent disclosure P: intermediate document			
T: theory or principle underlying the invention E: earlier patent document, but published on, or after, the filing date of the application D: document cited in the application L: document cited for other reasons B: member of the same patent family, corresponding document			

EP 0 823 479 A3

ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.

EP 97 11 3182

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

07-02-2000

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
EP 0655506	A	31-05-1995		US 5614365 A	25-03-1997	
				AT 143057 T	15-10-1996	
				AU 4193396 A	06-05-1996	
				CA 2201885 A	25-04-1996	
				CN 1170439 A	14-01-1998	
				DE 29513622 U	19-10-1995	
				DE 29513639 U	19-10-1995	
				DE 69400567 D	24-10-1996	
				DE 69400567 T	06-02-1997	
				DE 655506 T	28-09-1995	
				DK 655506 T	10-03-1997	
				EP 0727496 A	21-08-1996	
				ES 2072238 T	16-07-1995	
				FI 971611 A	13-06-1997	
				GR 95300040 T	31-07-1995	
				HU 77037 A	02-03-1998	
				JP 2691149 B	17-12-1997	
				JP 8205874 A	13-08-1996	
				WO 9612042 A	25-04-1997	
				ZA 9508761 A	13-05-1996	
WO 9610640	A	11-04-1996		US 5912155 A	15-06-1999	
				CA 2174944 A	11-04-1996	
				EP 0725827 A	14-08-1996	
				JP 9506783 T	08-07-1999	
				US 5939301 A	17-08-1999	



(19)

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 823 479 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:

11.02.1998 Bulletin 1998/07

(51) Int. Cl.⁶: C12N 15/54, C12N 9/12,

C12N 15/10, C12Q 1/68

(21) Application number: 97113182.6

(22) Date of filing: 31.07.1997

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LU MC
NL PT SE

• Kaiman, Lisa Vivian
San Francisco, CA 94131 (US)
• Reichert, Fred Lawrence
Oakland, CA 94610 (US)

(30) Priority: 06.08.1996 US 23376 P

(74) Representative:
Wächter, Dieter Ernst, Dr. et al
F.Hoffmann-La Roche AG
Patent Department (PLP),
124 Grenzacherstrasse
4070 Basel (CH)

(71) Applicant:
F. HOFFMANN-LA ROCHE AG
4070 Basel (CH)

(72) Inventors:
• Gelfand, David Harrow
Oakland, CA 94611 (US)

(54) Modified thermostable DNA polymerase

(57) The invention provides thermostable DNA polymerase enzymes that comprises the amino acid sequence SerGlnIleXaaLeuArgXaa (SEQ ID NO: 1), wherein "Xaa" at position 4 of this sequence is any amino acid residue but not a glutamic acid residue (Glu), preferably a glycine residue and "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile). The thermostable DNA polymerases of the invention have enhanced efficiency for incorporating unconventional nucleotides, such as ribonucleotides, into DNA products and are advantageous in many *in vitro* synthesis applications. Such enzymes are particularly useful for use in nucleic acid sequencing protocols and provide novel means for DNA sequence analysis with cost and efficiency advantages. Also claimed are nucleic acids encoding said polymerases, vectors and host cells comprising such a nucleic acid, as well as compositions for use in a DNA sequencing reaction, kits and methods for sequencing including such polymerases.

Description**Field of The Invention**

5 The present invention relates to thermostable DNA polymerases which have enhanced efficiency for incorporating ribonucleoside triphosphates. The invention provides methods and means for isolating such polymerases. The enzymes of the invention are useful for many applications and in particular for nucleic acid sequencing applications. Thus, the invention also provides improved methods for nucleic acid sequence analysis.

Background of the Invention

DNA sequencing generally involves the generation of four populations of single-stranded DNA fragments having one defined terminus and one variable terminus. The variable terminus generally terminates at specific nucleotide bases (either guanine (G), adenine (A), thymidine (T), or cytosine (C)). The four different sets of fragments are each 15 separated on the basis of their length. In one such procedure a high resolution polyacrylamide gel is used. Each band on such a gel corresponds to a specific nucleotide in the DNA sequence, thus identifying the positions in the sequence.

A frequently used DNA sequencing method is the dideoxy or chain-terminating sequencing method, which involves the enzymatic synthesis of a DNA strand (Sanger et al., 1977, Proc. Natl. Acad. Sci. USA 74:5463). Four separate syntheses are generally run, each reaction being caused to terminate at a specific base (G, A, T, or C) via incorporation of 20 an appropriate chain-terminating nucleotide, such as a dideoxynucleotide. The reaction products are easy to interpret since each lane corresponds only to either G, A, T, or C.

In the dideoxy chain-terminating method a short single-stranded primer is annealed to a single-stranded template. The primer is elongated at its 3' end by the incorporation of dideoxynucleotides (ddNTPs) until a dideoxynucleotide (ddNTP) is incorporated. When a ddNTP is incorporated, elongation ceases at that base. However, to assure fidelity of 25 DNA replication, DNA polymerases have a very strong bias for incorporation of their normal substrates, i.e. dNTPs, and against incorporation of nucleotide analogues, referred to as unconventional nucleotides. In the case of DNA synthesis, ribonucleotides (rNTPs) are considered unconventional nucleotides, because, like ddNTPs, rNTPs are not the normal *in vivo* substrate of a DNA polymerase. In the cell this property attenuates incorporation of abnormal bases such as deoxyinosine triphosphate (dITP) or rNTPs in a growing DNA strand.

30 Two frequently used automated sequencing methodologies are dye-primer and dye-terminator sequencing. These methods are suitable for use with fluorescent labeled moieties. Although sequencing can also be done using radioactive labeled moieties, fluorescence-based sequencing is increasingly preferred. Briefly, in dye-primer sequencing, a fluorescently labeled primer is used in combination with unlabeled ddNTPs. The procedure requires four synthesis reactions and up to four lanes on a gel for each template to be sequenced (one corresponding to each of the base-specific termination products). Following primer extension, the sequencing reaction mixtures containing dideoxynucleotide-incorporated termination products are routinely analyzed by electrophoresis on a DNA sequencing gel. Following separation by electrophoresis, the fluorescently-labeled products are excited with a laser at the bottom of the gel and the fluorescence is detected with an appropriate monitor. In automated systems, a detector scans the bottom of the gel during 35 electrophoresis, to detect whatever label moiety has been employed, as the reaction mixtures pass through the gel matrix (Smith et al., 1986, Nature 321:674-679). In a modification of this method, four primers are each labeled with a different fluorescent marker. After the four separate sequencing reactions are completed, the reaction mixtures are combined and the combined reaction mixtures are subjected to gel analysis in a single lane, whereby the different fluorescent tags (one corresponding to each of the four different base-specific termination products) are individually detected.

40 Alternatively, dye-terminator sequencing methods are employed. In this method, a DNA polymerase is used to incorporate dNTPs and fluorescently labeled ddNTPs onto the growing end of a DNA primer (Lee et al., 1992, Nucleic Acid Research 20:2471). This process offers the advantage of not having to synthesize dye-labeled primers. Furthermore, dye-terminator reactions are more convenient in that all four reactions can be performed in the same tube. Modified thermostable DNA polymerases having reduced discrimination against ddNTPs have been described (see

45 European Patent Application, Publication No. EP-A-655,506 and U.S. Patent Application Serial No. 08/448,223). An exemplary modified thermostable DNA polymerase is the mutated form of the DNA polymerase from *T. aquaticus* having a tyrosine residue at position 667 (instead of a phenylalanine residue), i.e. is a so called F667Y mutated form of Taq DNA polymerase. AmpliTaq[®] FS, manufactured by Hoffmann-La Roche and marketed through Perkin Elmer, reduces the amount of ddNTP required for efficient nucleic acid sequencing of a target by hundreds to thousands-fold. AmpliTaq[®] FS is a mutated form of the DNA polymerase from *T. aquaticus* having the F667Y mutation and additionally an aspartic acid residue at position 46 (instead of a glycine residue; G46D mutation).

50 There is a need for thermostable DNA polymerases that enable alternative nucleic acid synthesis methods for accurate and cost effective nucleic acid DNA sequence analysis. Fluorescence-based methods that do not require the

use of dideoxynucleotides would be desirable. The present invention addresses these needs.

Summary of the Invention

5 The present invention provides template-dependent thermostable DNA polymerase enzymes that comprises the amino acid sequence motif SerGlnIleXaaLeuArgXaa (SEQ ID NO: 1), whereby "Xaa" at position 4 of this sequence is any amino acid residue but not a glutamic acid residue (Glu) and "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile). As represented in the single letter code for amino acids this sequence motif can be represented as S Q I X L R V/I, wherein "X" at position 4 of this sequence is any amino acid residue but not a glutamic acid residue. Thermostable DNA polymerase enzymes having an amino acid sequence comprising the said sequence motif, wherein "X" at position 4 of this sequence motif is not a glutamic acid residue, have reduced discrimination against incorporation of ribonucleotides in comparison to previously known thermostable polymerases. In a growing DNA strand ribonucleotides are unconventional nucleotides. Thus, in a first aspect, the novel enzymes of the invention incorporate unconventional base analogues, such as ribonucleotides, into a growing DNA strand, several orders of magnitude more efficiently than previously identified thermostable DNA synthesizing enzymes. Genes encoding these enzymes are also provided by the present invention, as well as recombinant expression vectors and host cells comprising such vectors. With such transformed host cells large amounts of purified thermostable polymerase enzymes can be provided.

20 By the present invention a region or sequence motif within the amino acid sequence of thermostable DNA polymerases is identified which enhances the efficiency of the polymerase's ability to incorporate ribonucleotides while retaining the ability to faithfully incorporate deoxyribonucleotides. Alterations in this region, e.g. one or more amino acid exchanges (e.g. introduced by site specific mutagenesis) provides a thermostable polymerase enzyme which is capable of synthesizing an RNA or an RNA/DNA chimera or hybrid strand on a DNA template.

25 In another aspect, the invention provides improved methods and compositions for determining the sequence of a target nucleic acid, wherein the need for chain-terminating dNTPs is eliminated. By the improved methods provided herein, ribonucleotides (rNTPs) are incorporated into primer extension products. Because the subject enzymes accurately and efficiently incorporate either rNTPs or dNTPs, sequencing reactions can utilize mixtures of both nucleotides. Following primer extension, newly synthesized oligonucleotide products can be cleaved at the incorporated rNTPs by methods known in the art, e.g. by hydrolysis, thereby providing a population of fragments suitable for fractionation and 30 sequence analysis by conventional means, such as gel electrophoresis. These methods utilize the novel thermostable polymerase enzymes provided herein. Thus, in this aspect the invention provides thermostable DNA polymerase enzymes which are characterized in that the polymerase comprises the critical motif SerGlnIleXaaLeuArgXaa (SEQ. ID NO: 1), wherein "Xaa" at position 4 can be any amino acid residue but not a glutamic acid residue (Glu) and "Xaa" at position 7 is a valine residue (Val) or an isoleucine residue (Ile).

35 In another aspect of the invention, the modified polymerases described herein provide means for incorporating ribonucleotides or analogues containing a hydroxyl group, or other substitution, at the 2' position which, in comparison, is absent in conventional deoxyribonucleotides. These nucleotides can be differentially labeled, providing alternatives to the conventional use of dideoxynucleotides for DNA sequencing applications.

40 The mutant thermostable polymerase enzymes of the invention are characterized by the ability to more efficiently incorporate unconventional nucleotides, particularly ribonucleotides, than the corresponding wild-type enzymes. In a preferred embodiment of the invention, the unconventional nucleotide to be incorporated may be a chain-terminating base analogue, such as 2'-hydroxy - 3' deoxy ATP (cordycepin triphosphate) a "ribot terminator" analogue of ATP, or a non-chain-terminating nucleotide such as a rNTP.

45 In another aspect of the invention, mutant thermostable polymerase enzymes are provided which are characterized by the ability to more efficiently incorporate unconventional nucleotides, particularly ribonucleotides, than the corresponding wild-type enzymes. Thus, in this aspect the invention provides recombinant thermostable DNA polymerase enzymes which are each characterized in that (a) in its native form the polymerase comprises the amino acid sequence SerGlnIleGluLeuArgXaa (SEQ ID NO: 2), wherein "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile); (b) the amino acid sequence is mutated in the recombinant enzyme, preferably at position 4 of 50 this sequence so that the glutamic acid residue at position 4 is another amino acid residue, preferably a glycine residue; and (c) the recombinant enzyme has reduced discrimination against incorporation of ribonucleotides and ribonucleotide analogues in comparison to the native form of said enzyme.

55 In another aspect of the invention the polymerases of the invention provide a convenient means of fragmenting amplification products and primer extension products, such fragmented products may be useful in hybridization-based methodologies and a variety of sequence detection strategies.

The enzymes of the present invention and the genes encoding them can be used to provide compositions for use in DNA sequencing reactions that comprise a mixture of conventional nucleotides and at least one ribonucleotide or ribonucleotide analogue. In a preferred embodiment of the invention the unconventional nucleotide is a ribonucleotide,

and the ribonucleotide concentration is less than the concentration of the corresponding deoxyribonucleotide, i.e., the rNTP:dNTP ratio is 1:1 or less. The enzymes of the invention are also suitable for commercialization in kit formats, which kits may also include any of the following additional elements necessary for a nucleic acid sequencing reaction, such as e.g. dNTPs, rNTPs, buffers and/or primers.

5

Detailed Description of the Invention

The present invention provides novel and improved modified thermostable DNA polymerase enzymes, compositions and kits as defined in the appended set of claims. The enzymes of the invention more efficiently incorporate 10 unconventional nucleoside triphosphates than the previously known polymerases or the corresponding wild-type polymerase enzymes when from these novel polymerases are derived. DNA sequences encoding these modified enzymes, vectors for expressing the modified enzymes, and cells transferred with such vectors are also provided. The enzymes of the invention enable the practice of novel DNA sequencing methods which are advantageous over DNA sequencing procedures known from the prior art.

15 To facilitate understanding of the invention, a number of terms are defined below.

The term "conventional" when referring to nucleic acid bases, nucleoside triphosphates, or nucleotides refers to those which occur naturally in the polynucleotide being described (i.e., for DNA these are dATP, dGTP, dCTP and dTTP). Additionally, c7dGTP and dITP are frequently utilized in place of dGTP (although incorporated with lower efficiency) in vitro DNA synthesis reactions, such as sequencing. Collectively these may be referred to as deoxyribonucleoside triphosphates (dNTPs).

20 The term "expression system" refers to DNA sequences containing a desired coding sequence and control sequences in operable linkage, so that hosts transformed with these sequences are capable of producing the encoded proteins. To effect transformation, the expression system may be included in a vector, however, the relevant DNA may also be integrated into the host chromosome.

25 The term "gene" refers to a DNA sequence that comprises control and coding sequences necessary for the production of a recoverable bioactive polypeptide or precursor. The polypeptide can be encoded by a full-length gene sequence or by any portion of the coding sequence so long as the enzymatic activity is retained.

The term "host cell(s)" refers to both single cellular prokaryote and eukaryote organisms such as bacteria, yeast, and actinomycetes and single cells from higher order plants or animals when being grown in cell culture.

30 As used herein, the term "DNA sequencing reaction mixture" refers to a reaction mixture that comprises elements necessary for a DNA sequencing reaction. Thus, a DNA sequencing reaction mixture is suitable for use in a DNA sequencing method for determining the nucleic acid sequence of a target, although the reaction mixture may initially be incomplete, so that the initiation of the sequencing reaction is controlled by the user. In this manner, the reaction may be initiated once a final element, such as the enzyme, is added, to provide a complete DNA sequencing reaction mixture. Typically, a DNA sequencing reaction will contain a buffer, suitable for polymerization activity, nucleoside triphosphates and at least one unconventional nucleotide. The reaction mixture also may contain a primer suitable for 35 extension on a target by a polymerase enzyme, a polymerase and a target nucleic acid. Either the primer or one of the nucleotides is generally labeled with a detectable moiety such as a fluorescent label. Generally, the reaction is a mixture that comprises four conventional nucleotides and at least one unconventional nucleotide. In a preferred embodiment of the invention, the polymerase is a thermostable DNA polymerase and the unconventional nucleotide is a ribonucleotide.

40 The term "oligonucleotide" as used herein is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three, and usually more than ten. The exact size of an oligonucleotide will depend on many factors, including the ultimate function or use of the oligonucleotide.

45 Oligonucleotides can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphotriester method of Narang et al., 1979, *Meth. Enzymol.* 68:90-99; the phosphodiester method of Brown et al., 1979, *Meth. Enzymol.* 68:109-151; the diethylphosphoramidite method of Beaucage et al., 1981, *Tetrahedron Lett.* 22:1859-1862; the triester method of Matteucci et al., 1981, *J. Am. Chem. Soc.* 103:3185-3191; automated synthesis methods; or the solid support method of U.S. Patent No. 4,458,066.

50 The term "primer" as used herein refers to an oligonucleotide, whether natural or synthetic, which is capable of acting as a point of initiation of synthesis when placed under conditions in which primer extension is initiated. A primer is preferably a single-stranded oligodeoxyribonucleotide. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 35 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template for primer elongation to occur.

55 A primer can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ³²P, fluorescent dyes, electron-dense reagents, enzymes (as commonly used in ELISAs), biotin, or haptens and proteins for which antisera or monoclonal

antibodies are available.

The term "thermostable polymerase," refers to an enzyme which is stable to heat, is heat resistant and retains sufficient activity, to effect subsequent primer extension reactions when subjected to the elevated temperatures for the time necessary to effect denaturation of double-stranded nucleic acids. As used herein, a thermostable polymerase is suitable for use in a temperature cycling reaction such as the polymerase chain reaction (PCR). For a thermostable polymerase, enzymatic activity refers to the catalysis of the combination of the nucleotides in the proper manner to form primer extension products that are complementary to a template nucleic acid strand.

The heating conditions necessary for nucleic acid denaturation will depend, e.g., on the buffer salt concentration and the composition and length of the nucleic acids being denatured, but typically range from about 90°C to about 105°C, preferably 90°C to 100°C, for a time depending mainly on the temperature and the nucleic acid length, typically from a few seconds up to four minutes.

The term "unconventional" or "modified" when referring to a nucleic acid base, nucleoside triphosphate, or nucleotide, includes modification, derivations, or analogues of conventional bases, or nucleotides that naturally occur in DNA. More particularly, as used herein, unconventional nucleotides are modified at the 2' position of the ribose sugar in comparison to conventional dNTPs. Thus, although for RNA the naturally occurring nucleotides are ribonucleotides (i.e., ATP, GTP, CTP, UTP collectively rNTPs), because these nucleotides have a hydroxyl group at the 2' position of the sugar, which, by comparison is absent in dNTPs, as used herein, ribonucleotides are unconventional nucleotides. Ribonucleotide analogues containing substitutions at the 2' position, such as 2'-fluoro- or 2'-amino-substituted analogues, are within the scope of the invention. Additionally, ribonucleotide analogues may be modified at the 3' position, for example, by replacement of the normal hydroxyl group by a hydrogen group (3' deoxy), providing a ribonucleotide analogue terminator. Such nucleotides are also included within the scope of the term "unconventional nucleotides."

Since DNA is conventionally composed of dNTPs, incorporation of an rNTP would be unconventional and thus a rNTP would be an unconventional base. Consequently, in a preferred embodiment of the invention, for DNA primer extension methods including DNA sequencing methods, nucleic acid products contain both conventional and unconventional nucleotides, and predominantly comprise conventional nucleotides which are dNTPs.

Unconventional bases may be fluorescently labeled with, for example, fluorescein, or rhodamine; non-fluorescently labeled with, for example biotin; isotopically labeled with, for example, ^{32}P , ^{33}P , or ^{35}S ; or unlabeled.

In order to further facilitate understanding of the invention, specific thermostable DNA polymerase enzymes are referred to throughout the specification to exemplify the invention; however, these references are not intended to limit the scope of the invention. In a preferred embodiment the thermostable enzymes of the invention are utilized in a variety of nucleic acid sequencing methods, although the novel thermostable polymerases described herein may be used for any purpose in which such enzyme activity is necessary or desired. The enzyme can also be used in amplification reactions such as PCR.

The thermostable polymerases of the invention are characterized in that each contains the critical motif SerGlnIle-eXaaLeuArgXaa (SEQ ID NO: 1), whereby "Xaa" at position 4 of this sequence is any amino acid residue but not a glutamic acid residue (Glu) and "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile). Genes encoding thermostable polymerases which have a glutamic acid residue at the position 4 of the said motif can be modified as described herein to provide suitable modified polymerase enzymes. Said modified thermostable polymerase enzymes are characterized in that in comparison to the corresponding native or wild-type enzymes, they have a modification in the amino acid sequence motif SerGlnIle-eXaaLeuArgXaa (SEQ ID NO: 2), wherein "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile), i.e. said motif has been modified by a replacement of the glutamic acid residue at position 4 by another amino acid residue. The critical motif of a thermostable DNA polymerase provided by the present invention is shown below using the conventional three-letter amino acid code (Lehninger, Biochemistry, New York, New York, Worth Publishers Inc., 1970, page 67).

SEQ ID NO: 1 SerGlnIleXaaLeuArgXaa,

wherein "Xaa" at position 4 is any amino acid residue but is not a glutamic acid residue (Glu) and "Xaa" at position 7 is a valine residue (Val) or an isoleucine residue (Ile).

Both, gene sequences encoding and proteins containing this critical amino acid sequence, wherein Xaa at position 4 is not a glutamic acid residue (Glu), provide a polymerase having decreased discrimination against rNTPs, and are within the scope of the invention. Within the critical motif, additional modifications may be made with respect to other amino acid residues in this critical motif, preferably with respect to an amino acid residue selected from the group of glutamine (Gln or Q), leucine (Leu or L), or arginine (Arg or R).

The present invention is suitable for preparing thermostable DNA polymerase enzymes with advantageous properties by particular modification of the gene sequence encoding a thermostable DNA polymerase. In a preferred embodiment of the invention, the gene sequence and encoded enzyme are derived from a species of the genus *Thermus*, although non-*Thermus* eubacteria are included within the scope of the invention as described in detail below. Analogously, in view of the highly conserved nature of the now identified critical motif, novel thermostable DNA polymerases may be identified based upon their homology to, for example, Taq polymerase. Such thermostable polymerases are

within the scope of the present invention, as long as their amino acid sequence comprises the S Q I X L R V/I motif, wherein X is any amino acid residue but not glutamic acid residue and which amino acid sequence displays at least about 33 %, preferably at least about 60 %, more preferably at least about 80 % overall homology (sequence identity) in comparison to the amino acid sequence of the native Taq polymerase. The full-length sequence of said Taq polymerase is provided in WO 89/06691 and accessible under accession No. P90556 in the GENSEQ patent sequence data bank or under accession No. M26480 in the EMBL sequence data bank and under accession No. A33533 in the PIR sequence data bank.

Exemplary thermostable DNA polymerases of the present invention are recombinant derivatives of the native polymerases from the organisms listed in Table 1 below. Table 1 indicates the particular sequence of the critical motif and the position of the "X" residue for each of these native polymerases. Because each thermostable DNA polymerase is unique, the amino acid position of the critical motif is distinct for each enzyme. For those polymerases listed below, the amino acid residue in the "X" position of the critical S Q I X L R V/I motif is glutamic acid. The preferred polymerases of the present invention have a molecular weight in the range of 85'000 to 105'000, more preferably between 90'000 to 95'000. The amino acid sequence of these polymerases consists of about 750 to 950 amino acid residues, preferable between 800 and 850 amino acid residues. The polymerases of the present invention may also consist of about 540 or more amino acids and comprise at least the polymerase domain and a portion corresponding to the 3' to 5' exonuclease domain (the resulting polymerase may have 3' to 5' exonuclease activity or not) and possibly parts of the 5' to 3' exonuclease domain, which is contained on the first one-third of the amino acid sequence of many full-length thermostable polymerase enzymes.

For thermostable DNA polymerases not shown in Table 1, identifying the appropriate glutamic acid for modification is simple once the critical motif or consensus motif in the amino acid sequence is identified.

Regardless of the exact position within a thermostable DNA polymerase, the replacement of the glutamic acid (Glu) residue by another amino acid residue within the sequence motif SerGlnIleGluLeuArgXaa (SEQ ID NO: 2), wherein "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile) of the polymerase domain, 25 serves to provide thermostable polymerases having the ability to efficiently incorporate unconventional nucleotides. In a preferred embodiment, the glutamic acid is replaced by an amino acid having an uncharged polar R group such as glycine, serine, cysteine, threonine, or by an amino acid having a small nonpolar R group such as e.g. alanine. In a most preferred embodiment, the glutamic acid residue is replaced by a glycine residue (G). Amino acid and nucleic acid sequence alignment programs are readily available from the Genetics Computer Group, 575 Science Drive, Madison, 30 Wisconsin. Given the particular motif identified herein, these programs, including, for example, "GAP", "BESTFIT" and "PILEUP", serve to assist in the identification of the exact sequence region to be modified.

As it is evident from Table 1 below there are essentially two forms of the conserved sequence motif SerGlnIleGluLeuArgXaa (SEQ ID NO: 2) within the polymerase domain of thermostable DNA polymerase enzymes from thermophilic organisms. The sequence motif SerGlnIleGluLeuArgVal (SEQ ID NO: 3) is present in the native thermostable 35 polymerases from *Thermus* species such as e.g. from *Thermus aquaticus*, *Thermus caldophilus*, *Thermus thermophilus*, *Thermus flavus* and from *Thermus litoreus* as well as from the *Thermus* species sps17 and Z05. The sequence motif SerGlnIleGluLeuArgVal (SEQ ID NO: 3) is also present in the polymerase domain of other thermostable DNA polymerase enzymes, e.g. from *Thermosiphon africanus* and from various *Bacillus* strains such as *Bacillus caldotenax* and *Bacillus stearothermophilus*. The sequence motif SerGlnIleGluLeuArgIle (SEQ ID NO: 4) is e.g. present in 40 the native thermostable polymerases from *Thermotoga maritima*, *Thermotoga neapolitana* and *Anaerococcus thermophilum*.

45

50

55

Table 1

Organism	Amino Acid Consensus Motif	Position of Glutamic Acid
<i>Thermus aquaticus</i> (Taq)	S Q I X L R V/I	
<i>Thermus caldophilus</i> (Tca)	S Q I E L R V	615
<i>Thermus thermophilus</i> (Tth)	S Q I E L R V	617
<i>Thermus flavus</i> (Tfl)	S Q I E L R V	616
<i>Thermus filiformis</i> (Tfl)	S Q I E L R V	613
<i>Thermus</i> specie sps17	S Q I E L R V	613
<i>Thermus</i> specie Z05	S Q I E L R V	617
<i>Thermotoga maritima</i> (Tma)	S Q I E L R I	678
<i>Thermotoga neapolitana</i> (Tne)	S Q I E L R I	678
<i>Thermosiphon africanus</i> (Taf)	S Q I E L R V	677
<i>Anaerococcus thermophilum</i> (Ath)	S Q I E L R I	632
<i>Bacillus caldotenax</i> (Bca)	S Q I E L R V	659
<i>Bacillus stearothermophilus</i> (Bst)	S Q I E L R V	658, 661, or 736*

* depending on the amino acid sequence selected (see below)

The full nucleic acid and amino acid sequence for each of Taq, Tth, Z05, sps17, Tma, and Taf polymerases has been published in U.S. Patent No. 5,466,591 and are incorporated herein by reference. The sequences for the DNA polymerase from Tca, Tfl, Tne, Ath, Bca, and Bst have been published as follows: Tca in the EMBL sequence data bank under Accession No. U62584 (see also Kwon, 1997, *Mol. Cells* 7(2): 264-271); Tfl in Akhmetzhanov and Vakhitov, 1992, *Nucleic Acids Research* 20(21):5839; Tne in WO 97/09451 and WO 96/41014; Ath in the EMBL sequence data bank under Accession No. X98575 (for details on the Ath strain see Rainey et al., 1993, *J. Bacteriol.* 175 (15): 4772-2779; Bst in Uemori et al., 1993, *J. Biochem.* 113:401-410 and under EMBL sequence data bank Accession No. U23149 (see also Phang et al., 1995, *Gene* 163:65-68). Bst polymerase amino acid sequences comprising an E in the critical motif at position 658 are also provided by Japanese Patent publication JP 05/304 954A, European Patent publication No. EP-A-699,760, and Aliotta et al., 1996, *Genet. Anal.* 12: 185-195; the sequence is also available from the EMBL sequence data bank under Accession No. U33536. The sequence as published in *Gene* 163: 65-68 (1995), contains the "E" of the critical motif at residue number 661. Bst in Uemori et al., 1993, *J. Biochem.* 113:401-410 and under EMBL sequence data bank Accession No. D12982. The thermostable DNA polymerase from *Thermus filiformis* (see *FEMS Microbiol. Lett.* 22: 149-153, 1994; also available from ATCC Deposit No. 43280) can be recovered using the methods provided in U.S. Patent No. 4,889,818, as well as based on the sequence information provided in Table 1. Each of the above sequences and publications is incorporated herein by reference. The sequence homology (sequence identity) between the amino acid sequence of the native form of Taq polymerase as provided in WO 89/06691 and the Tfl polymerase mentioned above is 87.4 %. The corresponding homologies with respect to the Tth polymerase is 87.4 %, with respect to the Tca polymerase is 86.6 %, with respect to the Tca polymerase is 86.6 %, with respect to the Bst polymerase (Accession No. U23149) is 42.0 %, with respect to the Bca polymerase is 42.6 % and with respect to the Ath polymerase is 39.7 %.

As Table 1 demonstrates, the critical motif is remarkably conserved among the thermostable DNA polymerases. Where, "X" is a glutamic acid residue, alteration of the gene encoding the polymerase provides the enzyme of the invention, which readily incorporates rNTPs in comparison to, for example, Taq polymerase wherein the critical motif is not modified. Consequently, the invention relates to a class of enzymes which also includes, for example, the thermostable DNA polymerase, and corresponding gene and expression vectors from *Thermus osmimai* (Williams et al., 1996, *Int. J. Syst. Bacteriol.* 46 (2): 403-408); *Thermus sylvanus* and *Thermus chlorophilus* (Tenneiro et al., 1995, *Int. J. Syst. Bacteriol.* 45 (4): 633-639); *Thermus scotoductus* (Tenneiro et al., 1995, *Res. Microbiol.* 146 (4): 315-324); *Thermus brockianus* (Munster, 1986, *Gen. Microbiol.* 132: 1677) and *Thermus ruber*, Loginov et al., 1994 *Int. J. Syst. Bacteriol.* 34: 498-499; also available from ATCC Deposit No. 35948. Additionally, the invention includes, for example, the modified

forms of the thermostable DNA polymerases, and corresponding gene and expression vectors from *Thermotoga elfii* (Ravot et al., 1995, *Int. J. Syst. Bacteriol.*, 45: 312; also available from DSM Deposit No. 9442) and *Thermotoga thermarum* (Windberger et al., 1992, *Int. J. Syst. Bacteriol.*, 42: 327; also available from DSM Deposit No. 5069). Each of the above sequences and publications is incorporated herein by reference.

5 In a preferred embodiment of the invention, the critical motif to be modified is within the amino acid sequence LeuAspTyrSerGlnIleGluLeuArgValLeuAlaHisLeuSer (SEQ ID NO: 5). Thus, one aspect of the invention involves the generation of thermostable DNA polymerase mutants displaying greatly increased efficiency for incorporating unconventional nucleotides in a template-dependent manner. In a particularly preferred embodiment, the polymerase sequence comprises LeuAspTyrSerGlnIleGlyLeuArgValLeuAlaHisLeuSer (SEQ ID NO: 6). Such thermostable DNA polymerases are particularly suitable in processes such as DNA sequencing, DNA directed RNA synthesis, and *in vitro* synthesis of rNTP substituted DNA.

10 The production of thermostable DNA polymerases with enhanced efficiency for incorporating unconventional bases may be accomplished by processes such as site-directed mutagenesis. See, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, 1989, second edition, Chapter 15.51, "Oligonucleotide-Mediated

15 Mutagenesis", which is incorporated herein by reference. For example, a mutation of "A" to a "G" in the second position of the codon encoding glutamic acid at residue 615 in the *Thermus aquaticus* (Taq) DNA polymerase gene sequence (see SEQ ID NO: 7) results in more than a 500-fold increase in the efficiency of incorporation of unconventional nucleotides, as defined herein, while retaining the enzyme's ability to mediate PCR in the presence of conventional nucleotides, i.e., dNTPs. In Taq DNA polymerase this particular mutation results in an amino acid change of E (glutamic acid)

20 to G (glycine). Although this particular amino acid change significantly alters the ability of the enzyme to incorporate unconventional nucleotides, it is expected that the replacement of the glutamic acid residue by any other amino acid residues such as e.g. by a serine, cysteine, threonine, alanine, valine or leucine residue has the same effect. Other amino acid substitutions which replace E615 are therefore within the scope of the invention, although E615G represents a preferred embodiment. Thus, a critical aspect of the invention is that the fourth amino acid residue in the motif

25 of SEQ ID NO: 1 is not a glutamic acid residue.

Site-directed mutagenesis can also be accomplished by site-specific primer-directed mutagenesis. This technique is now standard in the art and is conducted using a synthetic oligonucleotide primer complementary to a single-stranded phage DNA to be mutagenized except for a limited mismatch representing the desired mutation. Briefly, the synthetic oligonucleotide is used as a primer to direct synthesis of a strand complementary to the plasmid or phage, 30 and the resulting double-stranded DNA is transformed into a phage-supporting host bacterium. The resulting bacteria can be assayed by, for example, DNA sequence analysis or probe hybridization to identify those plaques carrying the desired mutated gene sequence. Alternatively, "recombinant PCR" methods can be employed which are described in PCR Protocols, San Diego, Academic Press, Innis et al., editors, 1990, Chapter 22, entitled "Recombinant PCR" by Higuchi, pages 177-183.

35 As demonstrated in Table I, the glutamic acid within the critical motif of Taq polymerase is conserved in other thermostable DNA polymerases but may be located at a different but nearby position in the amino acid sequence. A mutation of the conserved glutamic acid within SEQ ID NO: 2 of *Thermus* species thermostable DNA polymerases and the related *Thermotoga*, *Thermosiphlo* and *Anaerococcus* species DNA polymerases, will have a similar enhancing effect on the ability of the polymerase to efficiently incorporate unconventional nucleotides in comparison to Taq polymerase 40 comprising SEQ ID NO: 2. Mutations of the glutamic acid residue within the critical motif in other thermostable DNA polymerases can be accomplished utilizing the principles and techniques used for site-directed mutagenesis. There are several sequence submissions for *Bacillus stearothermophilus* DNA polymerase in the GeneBank, or SwissProt/PIR databases. These sequences are highly related, but somewhat different from one another, but each contains the identical critical motif sequence SerGlnIleGluLeuArgXaa (SEQ ID NO: 2), wherein "Xaa" at position 7 of this sequence is a 45 valine residue (Val) or an isoleucine residue (Ile), although at different positions in the sequence.

Based on the publicly available amino acid and nucleic acid sequence information for thermostable DNA polymerases as described herein, it is also possible to construct, by conventional recombinant methodologies, chimeric polymerases which are composed of domains derived from different thermostable DNA polymerases. US Patent Nos. 5,466,591 and 5,374,553 describe methods for exchanging the various functional segments of thermostable polymerases, such as the 5' to 3' exonuclease domain, the 3' to 5' exonuclease domain and the polymerase domain to provide novel enzymes. The preferred chimeric thermostable polymerase enzymes comprise a 5' to 3' exonuclease domain, a 3' to 5' exonuclease domain and a polymerase domain, whereby one domain is derived from a different polymerase and whereby the polymerase domain comprises the critical motif sequence SerGlnIleXaaLeuArgXaa (SEQ ID NO: 1), wherein "Xaa" at position 4 of this sequence is any amino acid residue but not a glutamic acid residue (Glu) and "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile). Examples for such a chimeric molecules are Taq/Tma chimeric enzymes which are composed as specified in Table 2. As indicated in this Table the polymerase domain of these Taq/Tma chimeric enzymes contains the mutation in the critical motif specified above.

Table 2

	5' to 3' exonuclease domain	3' to 5' exonuclease domain	polymerase domain
5	Taq	aa. 1-289	aa. 290-422
Tma	aa. 1-291	aa. 292-484	aa. 485-893
10	Taq/Tma	aa. 1-289 (Taq)	aa. 292-484 (Tma)
Taq/Tma	aa. 1-289 (Taq)	aa. 292-484 (Tma)	aa. 485-893 (Tma) with E678G mutation

Plasmid pC1 has been deposited under the Budapest Treaty with the ATCC on July 17, 1996 and given Accession No. 98107. The plasmid pC1 contains a gene encoding a thermostable DNA polymerase that is mutated at the codon encoding the glutamic acid residue at position 615 of the amino acid sequence of native Taq polymerase, resulting in a mutated form of Taq polymerase having a glycine residue at position 613 (E615G mutated Taq polymerase having the sequence of SEQ ID NO: 8). This deposit provides alternative means for providing thermostable DNA polymerases having an enhanced efficiency for incorporating unconventional nucleotide analogues. Example 1 illustrates the use of flanking restriction sites suitable for subcloning the E615G mutation to create other thermostable DNA polymerase enzymes. Because the complete gene sequence for numerous thermostable DNA polymerases are known, other means for introducing a mutation at the codon encoding E 615, such as by restriction digestion and fragment replacement, or by site specific *in vitro* mutagenesis, are readily available to those of skill in the art based on the sequence information on the critical motif provided herein.

25 The modified gene or gene fragment prepared by site specific mutagenesis can be recovered from the plasmid, or phage by conventional means and ligated into an expression vector for subsequent culture and purification of the resulting enzyme. Numerous cloning and expression vectors, including mammalian and bacterial systems, are suitable for practicing the invention, and are described in, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, second edition, Cold Spring Harbor, 1989. For convenience, the present invention is exemplified utilizing the lambda derived PL promoter (Shimatake et al., 1961, *Nature* 222:128). Use of this promoter is specifically described in U.S. Patent Nos. 4,711,845 and 5,079,352.

30 The thermostable DNA polymerases of the present invention are generally purified from microorganisms such as e.g. *E. coli* which have been transformed with an expression vector operably linked to a gene encoding a wild-type or modified thermostable DNA polymerase. An example for a suitable host microorganisms is the *E. coli* strain DG116 described by Lawyer et al., 1993, *PCR Methods and Applications* 2:275-287, which strain is also available from the American Type Culture Collection under Accession No. ATCC 53601. Methods for purifying the thermostable polymerase are also described in, for example, Lawyer et al., 1993, *PCR Methods and Applications* 2:275-287.

35 Those of skill in the art will recognize that the above thermostable DNA polymerases, with enhanced efficiency for incorporating unconventional nucleotides, are most easily prepared by using methods of recombinant DNA technology. When one desires to produce one of the enzymes of the present invention, or a derivative or homologue of those enzymes, the production of a recombinant form of the enzyme typically involves the construction of an expression vector, the transformation of a host cell with the vector, and culture of the transformed host cell under conditions such that expression will occur. Means for preparing expression vectors, transforming and culturing transformed host cells are well known in the art and are described in detail in, for example, Sambrook et al., 1989, *supra*.

40 The present invention provides thermostable DNA polymerases suitable for use with ribonucleoside triphosphates for numerous applications including nucleic acid amplification, detection and DNA sequencing methods. The use of ribonucleotides in sequencing avoids the high cost of chain-terminating analogues, such as ddNTPs and importantly, facilitates the preparation of novel amplification products suitable not only for DNA sequence analysis but also other types of analysis such as electrophoresis or hybridization without the need to conduct subsequent DNA sequencing reactions.

45 Pyrophosphatase has been shown to enhance sequencing results using both mesophilic polymerases and thermostable DNA polymerase by decreasing the amount of pyrophosphorylation as extension products accumulate. Indeed, prior cycle sequencing methods require that the additional enzyme is included in the sequencing reaction. However, a very useful and advantageous aspect of the present invention is that pyrophosphatase is not required for DNA sequencing. Thus, use of the novel enzymes provided herein eliminates the need for the additional expense of adding a second enzyme into the sequencing reaction mixture.

50 By using the enzymes of the present invention, the amplification and sequencing reactions are combined, which saves time and materials, as well as simplifies the overall analysis. These advantages, and others, are available prima-

rily because the incorporation of both conventional nucleotides as well as ribonucleotides and ribonucleotide analogues into a primer extension product provides an RNA/DNA chimeric strand that is susceptible to hydrolysis of the RNA. The treatment does not affect the DNA backbone and provides a population of nucleic acid fragments each terminating at the position where a ribonucleotide was inserted in place of the corresponding dNTP. Hydrolysis is readily accomplished by various means including but not limited to alkali (e.g. by treatment with NaOH, e.g. at a final concentration of 0.2 M as shown in Example VI below), heat or enzymatic treatment with an RNase (Vogel et al., editors, *Informational Macromolecular*, New York, Academic Press, 1963, Chapter by Berg et al., Entitled "The Synthesis of Mixed Polynucleotides Containing Ribonucleotides and Deoxyribonucleotides by Purified Preparation of DNA Polymerase from *E. coli*", pages 467-483).

10 In a preferred embodiment, the present invention provides novel and improved compositions particularly useful for DNA sequencing methods. The novel enzymes described herein are advantageous in nucleic acid sequencing methods, using either dye-terminators or dye-primers, as well as other sequencing methods. As previously described, chain termination methods generally require template-dependent primer extension in the presence of chain-terminating nucleotides, resulting in a distribution of partial fragments which are subsequently separated by size. Standard dideoxy sequencing utilizes dideoxynucleoside triphosphates for chain termination and a DNA polymerase such as the Klenow fragment of *E. coli* Pol I (see Sanger et al., *supra*).

Thus, the basic dideoxy sequencing procedure involves (i) annealing an oligoribonucleotide primer to a template; (ii) extending the primer with DNA polymerase in four separate reactions, each containing one labeled nucleotide, or a labeled primer, a mixture of unlabeled dNTPs, and one chain-terminating ddNTP; (iii) resolving the four sets of reaction products by means of, for example, high-resolution denaturing polyacrylamide/urea gel electrophoresis, capillary separation or by other resolving means; and (iv) producing an autoradiographic image of the gel that can be examined to infer the sequence. Alternatively, mass spectrometry methods or hybridization-based methods, using fluorescently labeled primers or nucleotides, can be used to derive DNA sequence information.

The availability of thermostable polymerases, such as Taq polymerase, has resulted in improved methods for sequencing (see U.S. Patent No. 5,075,216) and modifications thereof referred to as "cycle sequencing." In cycle sequencing, cycles of heating and cooling are repeated allowing numerous extension products to be generated from each target molecule (Murray, 1989, *Nucleic Acids Research* 17:8889). The asymmetric amplification of target sequences complementary to the template sequence, in the presence of dideoxy chain terminators, produces a family of extension products of all possible lengths.

30 Following denaturation of the extension reaction product from the DNA template, multiple cycles of primer annealing and primer extension occur in the presence of dideoxy terminators. Thermostable DNA polymerases have several advantages in cycle sequencing; they tolerate the stringent annealing temperatures which are required for specific hybridization of primer to nucleic acid targets as well as tolerating the multiple cycles of high temperature denaturation which occur in each cycle, i.e., 90-95°C. For this reason, various forms of AmpliTaq® DNA polymerase have been included in Taq cycle sequencing kits commercialized by Perkin Elmer, Norwalk, CT.

Nevertheless, the property of Taq DNA polymerase, to discriminate against incorporation of unconventional nucleotides, such as ddNTPs, presents a problem when it is used for cycle sequencing, where ddNTPs or fluorescently labeled ddNTPs must be incorporated as chain terminators. Generally, prior to the present invention, DNA sequencing with thermostable DNA polymerases required a mixture of chain-terminating nucleotides, generally dideoxynucleotides, at high concentrations, to insure that a population of extension products would be generated representing all possible fragment lengths over a distance of several hundred bases. Frequently, to address this cost issue, protocols utilized very low concentrations of conventional dNTPs, making the reactions inefficient. These reaction mixtures, having a low dNTP concentration and a high ddNTP concentration, create an environment wherein the thermostable polymerase is essentially starved for nucleotide substrates.

45 Even with the advent of modified enzymes, such as AmpliTaq® DNA polymerase FS which allow the concentration of dNTPs to be increased to more optimal levels, the prior enzymes still rely on the presence of the costly ddNTPs for DNA sequencing. In contrast, the present invention provides enzymes that not only allow the concentration of dNTPs to be increased, but avoid the use of the costly ddNTPs by using instead dNTPs for incorporation into the growing strand. The ability of novel enzymes to efficiently effect partial ribonucleotide substitution facilitates the generation of DNA sequencing ladders in the absence of a separate reaction for incorporating a terminating nucleotide.

The choice of unconventional nucleotide analogues suitable for use in DNA sequencing methods was previously dictated by the ability of the thermostable DNA polymerase to incorporate said analogues. Unfortunately said nucleotide analogues are rather expensive. For example, the costs of ddNTPs is approximately 25X greater than the cost of either rNTPs or dNTPs. Because prior thermostable DNA polymerases were unable to efficiently incorporate rNTPs in a template directed manner into a growing DNA strand, such ribonucleotides, which are readily available and inexpensive, were not an option for use in DNA sequencing with a thermostable DNA polymerase. The present invention eliminates the need for ddNTPs in DNA sequencing reactions. Thus, in one aspect the invention provides methods for DNA sequencing analysis that are significantly less expensive than prior chain termination methods.

The presence of manganese in a primer extension reaction can influence the ability of a polymerase to accurately insert the correctly based paired nucleotide. Manganese can be used to force incorrect base pairing or to ease the discrimination against insertion of a nucleotide analogue. Manganese has been used by researchers to induce mutagenesis in DNA replication or amplification procedures. Thus, manganese can affect the fidelity of a polymerization reaction, as well as the yield of a reaction. The resulting sequence may be incorrect or, in a DNA sequencing method, the resulting information may be ambiguous. The present methods do not require that manganese is included as the divalent cation in the sequencing reaction mixture to force the polymerase to insert an unconventional nucleotide. In contrast to prior DNA polymerases, the present invention identifies the critical motif within the polymerase domain for controlling the enzyme's ability to discriminate between 2' substituted and unsubstituted nucleotides without the need for manganese.

The enzymes of the present invention do not require high concentrations of the unconventional base analogues for sequencing. Prior to the present invention unconventional base analogues and the corresponding conventional bases were generally present at a ratio (e.g., dATP:dATP) ranging from approximately 1.3:1 to 24:1 for chain termination DNA sequencing methods (see also U.S. Patent No. 5,075,216 of Ilnis et al.). In comparison, the thermostable polymerases provided by the present invention allow the ratio of unconventional base analogues to conventional bases to be reduced from a hundred to several thousand fold. A rNTP:dNTP ratio of 1:1 or less, in combination with the novel enzymes provided herein, is sufficient for DNA sequence analysis. In a preferred embodiment of the invention, the rNTP:dNTP ratio is reduced to less than 1:8. The ratio of 2' substituted nucleotide to the corresponding natural dNTP may be as low as 1:80 or 1:200, depending on the particular experimental design and desired length of fragments.

Thus, because the present enzymes readily incorporate unconventional nucleotides, such as 2' substituted nucleotides, it is not necessary to force incorporation of the rNTP by using a high concentration of rNTP and a limiting concentration of the corresponding dNTP. Accordingly, the present methods enable the use of optimal concentrations of dNTPs in combination with low amounts of rNTPs.

When modified polymerase enzymes in accordance with the present invention are used in a suitable sequencing method, such as e.g. dye-primer sequencing, good DNA sequencing results are obtained with a dNTP concentration in the range of 50-500 μ M of each dNTP. Preferably the dNTP concentration is between 100-300 μ M. In these ranges the corresponding rNTP may be present at about the same concentration as the dNTP, or less. Preferably the rNTP is present at about 0.1 μ M-100 μ M, most preferable the rNTP is present at about 2.5 μ M to 25 μ M.

The concentration of rNTPs suitable for use with the present modified enzymes can be readily determined by titration and optimization experiments by those of ordinary skill in the art. The amount of rNTP or analogue needed will be affected by the type of experiment and may be influenced by the target size and purity as well as the choice of buffer and the particular species of enzyme.

The ratio of rNTP:dNTP will determine the frequency with which rNTPs are inserted into the growing oligonucleotide. Because hydrolysis will occur at each incorporated rNTP, the ratio of rNTP:dNTP can be adjusted to provide the user with flexibility to increase or decrease the size of the resulting fragments.

As is well understood, DNA is a polymer synthesized from dNTPs. Each deoxynucleoside triphosphate comprises a ribose sugar which contains a hydroxyl group at the 2' position and a hydrogen at the 2' position. Ribonucleotides also contain a hydroxyl group at the 3' portion of the sugar. However, rNTPs are distinguished from dNTPs at the 2' position of the sugar, where a second hydroxyl group replaces the hydrogen atom. In the present context, rNTPs exemplify the ability of the enzymes of the present invention to accurately incorporate 2' substituted nucleotides. However, the compounds of the invention are not limited to the use of unconventional nucleotides which are ribonucleotides. Modification of the thermostable polymerase sequence at the critical domain identified herein enables template directed incorporation of alternative 2' substituted nucleotides, such as 2'-hydroxyl, 3'-deoxy nucleotides and substituted 2'-fluoro or amino nucleotides.

As is described in the examples herein, the incorporation of 3'-deoxy, 2'-hydroxy ATP, referred to herein as cordycepin triphosphate, is facilitated by the presence of a second mutation in the thermostable polymerase which reduces discrimination against incorporation of a nucleotide containing a deoxy at the 3' position of the ribose. Such enzymes have been previously described for example in EP-A-655506 and in U.S. Serial No. 08/448,223, filed May 23, 1995, which are incorporated herein by reference. ATCC Deposit No. 69820, deposited under the Budapest Treaty on May 10, 1995, provides the gene encoding a modified thermostable DNA polymerase of *Thermus aquaticus* that has reduced discrimination against incorporating analogues such as ddNTPs. Dideoxynucleotides have a substituted 3' position in comparison to conventional dNTPs. Thus, in combination with the present invention, the double mutation, exemplified herein by a E615G, F667Y *Tag* polymerase mutant, provides means for utilizing nucleotide analogues which are substituted at the 3' and 2' positions of the ribose, in comparison to dNTPs (see Examples III and V).

A particular application of the invention is a rNTP sequencing method, wherein the sequencing primer is detectably labeled with a distinguishable fluorescent or radioactive tag. Unlike dNTPs, incorporation of an unmodified rNTP does not result in a chain termination event. The DNA sequencing reaction comprising both rNTPs and dNTPs in combination with an enzyme of the invention, produces a mixture of randomly substituted primer extension products susceptible

to cleavage at the 3'-5' phosphodiester linkage between a ribo- and an adjacent deoxyribonucleotide. Following primer extension in, for example, PCR amplification or cycle sequencing, and prior to resolving the primer extension products, by, for example, gel electrophoresis, the reaction mix is treated with either alkali, heat, a ribonuclease or other means for hydrolyzing the extension products at each occurrence of a ribonucleotide. For each labeled primer extension product, only the most 5' fragment, which is the immediate extension product of the labeled primer, is detectable on a sequencing gel. For a given target, analysis of the resulting sequencing gel provides a sequencing ladder, i.e., series of identifiable signals in the G, A, T, and C, lanes corresponding to the nucleic acid sequence of the target. The resulting sequencing ladder provides the same information whether the method utilizes ddNTPs by conventional means, or rNTPs and the novel thermostable polymerases described herein. Thus, by use of the present invention, expensive ddNTPs are no longer required for DNA sequencing (see Example V).

In an alternative sequencing method, chain-terminating ribonucleotides are employed. In this embodiment of the invention, 2'-hydroxy, 3'-deoxy nucleotides, such as cordycepin triphosphate, are utilized as terminators. These rNTP analogues can be fluorescently labeled and utilized for DNA sequencing. Lee *et al.* (*supra*) have described the use of dye-terminator ddNTPs. EP-A-655,506 and U.S. Serial No. 08/448,223, filed May 23, 1995 describe modified enzymes for use with ddNTPs. A thermostable DNA polymerase comprising both the modification present in AmpliTaq[®] DNA polymerase FS (see above) and those specified in SEQ ID No: 1, wherein X is not glutamic acid (E), as described herein, can be used for efficiently incorporating the labeled rNTP analogues in a chain termination sequencing reaction. This process may be automated and does not require synthesis of dye labeled primers. Furthermore, because dye-terminator reactions allow all four reactions to be performed in the same tube, they are more convenient than dye-primer methods. The 2'-hydroxy, 3'-deoxy nucleotides can be synthesized from commercially available 3' nucleotides (3' dA, 3' dC, 3' dG and 3' dT, e.g. available from Sigma Chemical Corporation, St. Louis, MO) and adding a 5' triphosphate as described in Ludwig, *Bioephosphates and Their Synthesis Structure, Metabolism and Activity*, editors, Bruylants and Stec, Amsterdam, Elsevier Science Publishers, 1987, pages 201-204.

In addition to the utility of the enzymes of the present invention in novel sequencing methods, the modified enzymes described herein are useful in a number of molecular biology applications. In one embodiment, the modified enzyme is used in an amplification reaction mixture comprising both conventional and unconventional nucleotides, for example, dNTPs and at least one detectably labeled rNTP; the labels which include, for example, fluorescent labels or radioisotopes. Template directed synthesis of a complementary strand provides a DNA product containing ribonucleoside monophosphates at various positions along its length. Heat and/or alkali treatment hydrolyzes the nucleic acid extension product at each ribonucleotide. Thus, a family of DNA segments is provided wherein each fragment contains one label moiety at its 3' end. The size of the resulting nucleic acid fragments can be modified by adjusting the ratio and amount of rNTP included in the reaction.

The amplification of a target using rNTPs and the present enzymes provides numerous advantages depending upon the particular application. In the method described above using a labeled rNTP, the resulting family of fragments are all labeled with equal intensity: one label per oligonucleotide fragment. Procedures such as nucleic acid detection using an oligonucleotide probe array fixed to a silicon chip, optimally require that the amplified target is randomly fragmented within a fixed reproducible size range to limit formation of secondary structures for controlling hybridization kinetics. Further, for detecting hybridization to an array of thousands of probes on a chip, it may be preferable that the nucleic acid fragments are labeled with equal intensity. The present invention provides a means for producing families of fragments that meet this standard, and thereby facilitates the use of alternative detection formats such as the chip-based methods described by, for example, Cronin *et al.*, 1996, *Human Mutation* 7:244-255.

In another embodiment, the use of one labeled primer in an amplification reaction which comprises a thermostable polymerase of the invention and both rNTPs and dNTPs provides a means of simultaneously performing amplification and sequencing reactions. This method requires that four separate amplification reactions are conducted, one for each rNTP. Thus, for example, because the enzyme of the invention is suitable for target amplification by, for example, PCR, or other amplification means, the resulting product, if it is present, can be detected by conventional methods such as gel electrophoresis or probe hybridization using a portion of the reaction product. These detection methods will not result in hydrolysis of the incorporated ribonucleotides, and the RNA/DNA chimeric strands will behave as expected for a conventional nucleic acid amplification product. If a desired product is detected, a remaining portion of the same reaction mixture can be treated with alkali and analyzed by gel electrophoresis for nucleic acid sequence determination. Thus, following detection of the product, a subsequent sequencing reaction is unnecessary. This simplified procedure saves time and materials and provides increased accuracy by removing steps: the detected product is the sequenced product.

A similar procedure with four labeled rNTPs and one biotinylated primer could also be used. After amplification, the product is cleaved with alkali and the primer associated products are removed by reaction with streptavidin coated beads. The captured products are subsequently analyzed on a sequencing gel. This modification allows the sequencing reaction to be done in one tube, thus eliminating the need for four separate amplifications.

In another aspect of the invention, the enzymes described herein are useful for preparing RNA from a DNA tem-

plate or for making substituted DNA for alkali mediated sterilization without the use of conventional sterilizing agents such as uracil-N-glycosylase (UNG), as described in International Patent publication No. WO 92/01814.

In an exemplified embodiment, the thermostable polymerase also contains a mutation in the 5'-3' exonuclease domain that serves to greatly attenuate this exonuclease activity. Modified forms of Taq polymerase are described in U.S. Patent No. 5,466,591. In one embodiment of that invention, the codon encoding the glycine (G) residue at amino acid position 46 has been replaced with a codon encoding aspartic acid (D). The resulting enzyme has enhanced utility in cycle sequencing reactions due to the decreased 5'-3' exonuclease activity and is a preferred background for use with the present invention. The polymerase domain amino acid sequence and polymerase activity are unaffected by the presence of the (G46D) mutant in comparison to the wild-type enzyme.

In a commercial embodiment of the invention, kits for nucleic acid sequencing comprising a thermostable polymerase in accordance with the present invention represent a commercial embodiment of the invention. Such kits typically include additional reagents for DNA sequencing such as e.g. rNTPs, dNTPs, and appropriate buffers. Where rNTPs are unlabeled, a labeled primer may also be included.

The following examples are offered by way of illustration only and are by no means intended to limit the scope of the claimed invention.

Example I

Expression of a Modified Taq Polymerase Gene Having Reduced Discrimination Against Unconventional Nucleotides

The C-terminal amino acid portion of Taq DNA polymerase encodes the polymerase active site domain (Lawyer et al., 1993, *PCR Methods and Applications* 2:275-287, which is incorporated herein by reference). A DNA fragment containing this region was isolated from the full-length Taq gene and mutagenized by PCR amplification in the presence of manganese (Leung et al., 1989, *Technique* 1(1):11-15). For this example, all restriction enzymes were purchased from New England Biolabs, Beverly, MA. The mutagenized fragments were digested with *Pst*I and *Bgl*II and cloned into a Taq expression plasmid, here the plasmid pLK102, which had been digested with *Pst*I and *Bgl*II. Plasmid pLK102 is a modified form of Taq expression plasmid pSYC178 (Lawyer et al., *supra*). The *Hinc*II/*Eco*RV fragment located 3' to the polymerase coding region was deleted to create plasmid pLK101. A 898 base pair *Pst*I-*Bgl*II fragment was subsequently deleted from pLK101 and replaced by a short *Pst*I-*Eco*RV-*Bgl*II oligonucleotide duplex to create plasmid pLK102. Thus, this deletion removes 900 base pair from the 3' end of the Taq DNA pol gene and replaces it with a short piece of DNA.

The resulting expression plasmids were transformed into *E. coli* strain N1624 (described by Gottesman, 1973, *J. Mol. Biol.* 77: 531; also available from the *E. coli* Genetic Stock Center at Yale University, under strain No. CGSC #5066) and the resulting transformants were screened for the ability to efficiently incorporate rNTPs in comparison to the wild-type enzyme. Using this procedure, mutant C1 was identified as having the ability to more efficiently incorporate rNTPs.

To determine which portion of the Taq polymerase gene was responsible for the altered phenotype, the mutagenized Taq expression plasmid, named pC1, isolated from mutant C1, was digested with various restriction enzymes and the resulting restriction fragments were subcloned into the wild-type Taq DNA polymerase gene of pLK101, replacing the unmutagenized restriction fragments. Analysis of the resulting subclones indicated that the mutation responsible for the phenotype was contained within a 265 base pair *Nhe*I to *Bam*H I restriction fragment.

DNA sequence analysis was performed on this region of pC1 using the ABI PRISM® Dye Terminator Cycle Sequencing Core Kit with AmpliTaq® DNA polymerase FS from Applied Biosystems, Foster City, CA, and the Applied Biosystems Model 373A DNA Sequencing System. The sequence analysis identified two missense mutations in the Taq polymerase gene between the *Nhe*I and *Bam*H I sites. A mutation at amino acid position 615 caused a glutamic acid (E) to be replaced by a glycine residue (G) and another mutation at position 653 replaced an alanine (A) residue with a threonine (T). Numbering is initiated at the codon encoding the first methionine residue of the mature protein, as in U.S. Patent No. 5,079,352. The E615G mutation was caused by a GAG to GGG change in codon 615. The A653T mutation was caused by a GCC to ACC change at codon 653. Plasmid C1 in *E. coli* host strain N1624 was deposited under the Budapest Treaty with the ATCC on July 17, 1996, and given accession No. 98107.

The two point mutations were separately analyzed by subcloning each separately into a wild-type Taq polymerase gene, using recombinant PCR (Innis et al. editors, *PCR Protocols*, San Diego, Academic Press, 1990, Chapter 22, Entitled "Recombinant PCR", Higuchi, pages 177-183). The resulting expression products were analyzed to determine whether E615G or A653T or both mutations were responsible for the ribonucleotide incorporation phenotype. The results of this experiment indicated that the E615G mutation was solely responsible for the mutant phenotype.

For further analysis and quantitation of the incorporation efficiency of nucleotide analogues, the 265 base pair *Bam*H-*Nhe*I PCR fragment containing E615G was cloned into a Taq expression vector, pRDA3-2. Expression vector pRDA3-2 contains the full-length Taq gene operably linked to the phage lambda PL promoter. The exonuclease domain of the Taq gene in this vector contains a point mutation at the codon encoding glycine, amino acid residue 46, that

reduces 5'-3' exonuclease activity. However, the gene sequence within the polymerase domain of the expression vector pRDA3-2 is identical to the wild-type Taq gene sequence. Plasmid RDA3-2 is fully described in U.S. Patent No. 5,466,591, which is incorporated herein by reference, wherein the plasmid is referred to as "clone 3-2." Plasmid pRDA3-2 was digested with *Bam*H and *Nhe*I and the 265 base pair PCR fragment was ligated into the vector by conventional means.

5 The resulting plasmid, pLK108, was transformed into *E. coli* strain DG116 (Lawyer et al., 1993, *supra*, also available from the American Type Culture Collection under ATCC No. 53606). The plasmid pLK108 encodes a thermostable DNA polymerase herein referred to as G46D, E615G Taq. A mutant, G46D, E615G, F667Y Taq, was created by combining the E615G and F667Y mutations by recombinant PCR into a *Bam*H-*Nhe*I fragment. This fragment was cloned 10 into plasmid pRDA3-2 to create plasmid pLK109. The expressed thermostable DNA polymerase protein from plasmids pLK108 and pLK109 were purified according to the method described by Lawyer et al., 1993, *supra*, although the chromatography steps were omitted. The sequence of the insert was confirmed by DNA sequence analysis. An additional mutation in the sequence was detected in the pLK108 insert; however, this mutation does not change the amino acid sequence of the protein.

15 Following partial purification, the activity of the modified enzyme was determined by the activity assay described in Lawyer et al., 1989, *J. Biol. Chem.* 264:6427-6437, which is incorporated herein by reference. The activity of the modified enzyme was calculated as follows: one unit of enzyme corresponds to 10 nmoles of product synthesized in 30 minutes. DNA polymerase activity of the wild-type enzyme is linearly proportional to enzyme concentration up to 80-100 pmoles dCMP incorporated (diluted enzyme at 0.12-0.15 units per reaction). Activity of the E615G, G46D and E615G, 20 F667Y, G46D mutants is linearly proportional to enzyme concentrations up to 0.25-3 pmoles dCMP incorporated (diluted enzyme at 6 x 10⁻⁴ to 5 x 10⁻³ units per reaction). This enzyme preparation was utilized in the incorporation and sequencing reactions described in Examples III-V. For Examples II and VI, enzyme was purified as described in Lawyer et al. (*supra*).

25 Example II

Assay to Compare Efficiency of Incorporation

The relative activity G46D and G46D, E615G Taq to incorporate rNTPs was determined by measuring the amount 30 of [α -³²P]rNTP each enzyme could incorporate at limiting enzyme concentration into an activated salmon sperm DNA template. To measure the incorporation of rATP, a reaction mixture was prepared so that the final concentrations in a 50 μ l reaction were: 12.5 μ g activated salmon sperm DNA, prepared as described below, 200 μ M each dCTP, dGTP and dTTP (Perkin Elmer, Norwalk, CT), 100 μ M [α -³²P]rATP, 1 mM β -mercaptoethanol, 25 mM N-tris(hydroxymethyl)methyl-3-amino-propanesulfonic acid (TAPS) pH 9.5, 20°C, 50 mM KCl and 2.25 mM MgCl₂.

35 Similar assay mixtures were prepared to measure the incorporation of rCTP, rGTP and rUTP. In each case, the rNTP was radiolabeled and present at 100 μ M and the three remaining dNTPs (dATP, dGTP and dTTP for rCTP, dATP, dCTP and dTTP for rGTP and dATP, dCTP and dGTP for rUTP) were present at 200 μ M each. As a standard, incorporation of the corresponding [α -³²P]dNTP by each enzyme was also measured. The assay mixture for these assays was similar to the rNTP incorporation assay above except that each [α -³²P]rNTP was replaced with 100 μ M of the corresponding [α -³²P]dNTP. Crude salmon sperm DNA, 1g/l, from Worthington Biochemical, (Freehold, NJ) was activated 40 by incubation in 10 mM Tris-HCl, pH 7.2, 5 mM MgCl₂, at 2°C-6°C for 96 hours. EDTA and NaCl were then added to 12.5 mM and 0.1 M, respectively. The DNA was then extracted with phenol/chloroform and then ethanol precipitated and resuspended in 10 mM Tris, 1 mM EDTA, pH 7.5. The activated DNA preparation was then dialyzed against the same buffer.

45 Forty-five microliters of each reaction mixture were aliquoted into five 0.5 ml tubes (e.g. Eppendorf) for each of the 5'-labeled nucleotide precursors. Thus, each of G46D Taq and G46D, E615G Taq were assayed in duplicate with one tube remaining for a negative control. The polymerization reaction in two tubes of each assay mix was initiated with 5 μ l of either G46D Taq polymerase (0.02 units) or G46D, E615G Taq (0.002 units). As a control for the level of background, 5 μ l of enzyme dilution buffer rather than enzyme was added to the negative control reaction.

50 Each reaction was vortexed briefly and incubated for 10 minutes at 75°C. The reactions were stopped by addition of 10 μ l 60 mM EDTA and stored on ice. For each sample, 50 μ l aliquots of the 60 μ l reaction were diluted with 1 ml 2 mM EDTA, 5 μ g/ml sheared salmon sperm DNA. The DNA was precipitated with TCA using standard procedures and collected on GF/C filter discs (Whatman, Kent, England). The amount of incorporated [α -³²P]labeled nucleotide or ribonucleotide was quantitated by liquid scintillation spectrometry and the number of pmoles incorporated was then calculated. The number of pmoles of each NTP incorporated by each enzyme was normalized to the number of pmoles of the corresponding [α -³²P]dNTP incorporated by each enzyme. The resulting data is shown below.

Incorporation Ratio of rNTP to dNTP for G46D and G46D, E615G Taq								
Enzyme	pMoles Incorporated (percent)							
	dATP	rATP	dCTP	rCTP	dGTP	rGTP	dTTP	rUTP
G46D	27.74 (100%)	0.052 (0.18%)	34.6 (100%)	0.76 (0.22%)	36.94 (100%)	0.133 (0.36%)	28.79 (100%)	0 (0)
G46D, E615G	0.67 (100%)	1.41 (210%)	2.82 (100%)	5.33 (189%)	3.27 (100%)	5.96 (181%)	0.688 (100%)	0.545 (79%)

These results indicate that G46D, E615G incorporates ribonucleotides more than 500-fold more efficiently than can G46D (e.g. for rGTP 181: 0.36 = 502-fold, for rCTP 189: 0.22 = 859-fold and for rATP 210: 0.18 = 1166-fold more efficient). Thus, a missense mutation in the polymerase gene at codon 615, provided a novel phenotype: a thermostable DNA polymerase capable of efficiently incorporating ribonucleotides in addition to deoxyribonucleotides.

Example III

Assay to Compare Efficiency of Incorporation of 3'deoxy ATP (Cordycepin)

The relative ability G46D; G46D, E615G; G46D, E615G, F667Y and G46D, F667Y Taq to incorporate 3'-deoxy adenosine 5'-triphosphate (cordycepin triphosphate) was determined by measuring the amount of [α -³²P]cordycepin triphosphate each enzyme could incorporate at limiting enzyme concentration into an activated salmon sperm DNA template. To measure the incorporation of [α -³²P]cordycepin triphosphate, the assay was composed so that the final concentrations in a 5 μ l reaction were: 12.5 μ g activated salmon sperm DNA, 200 μ M each dCTP, dGTP and dTTP, 50 μ M dATP (Perkin Elmer), 50 μ M [α -³²P]-3'dATP/3' dATP (New England Nuclear, Sigma), 1 mM β -mercaptoethanol, 25 mM N-tris(hydroxymethyl)methyl-3-amino-propanesulfonic acid (TAPS) pH 9.5, 20°C, 55 mM KCl and 2.25 mM MgCl₂.

Forty-five microliters of each reaction mixture were aliquoted into nine 0.5 ml tubes, thus each reaction will be done with either G46D; G46D, E615G; G46D, E615G, F667Y or G46D, F667Y Taq in duplicate with one tube remaining for a no enzyme control. The polymerization reaction in two tubes of assay mix was started with 5 μ l (0.058 units) of G46D Taq polymerase. The same was done for G46D, E615G Taq (0.0025 units), G46D, E615G, F667Y Taq (0.0034 units) or G46D, F667Y Taq (0.083 units). As a control for the level of background, the one remaining tube was started with enzyme dilution buffer rather than enzyme.

Each reaction was vortexed briefly and incubated for 10 minutes at 75°C. The reactions were stopped by addition of 10 μ l 60 mM EDTA and stored on ice. For each sample, 50 μ l aliquots of the 60 μ l reaction were diluted with 1 ml 2 mM EDTA, 50 μ g/ml sheared salmon sperm DNA. The DNA was precipitated with TCA using standard procedures and collected on GF/C filter discs (Whatman, Kent, England). The amount of incorporated [α -³²P] labeled nucleotide was quantitated by liquid scintillation spectrometry and the number of pmoles incorporated was then calculated. The number of pmoles of [α -³²P]cordycepin triphosphate incorporated by each enzyme was divided by the number of units of each enzyme used in the assay to give the pmoles incorporated per unit enzyme. A chart of this data is shown below.

Incorporation of [α - 32 P]cordycepin by G46D; G46D, E615G; G46D, E615G, F667Y and G46D, F667Y Taq		
	Enzyme	pmoles Incorporated per unit of Enzyme
5	G46D	0.221
10	G46D, E615G	1.56
	G46D, E615G, F667Y	893.6
	G46D, F667Y	0.74

15 These results indicate that both the E615G and the F667Y mutations are required for the efficient incorporation of the cordycepin molecule into DNA.

20 Example IV

Alkaline Cleavage DNA Sequencing Using G46D, E615G Taq DNA Polymerase

This example demonstrates the application of the modified polymerase of the invention to alkaline cleavage sequencing, utilizing partially rNTP substituted DNA. The ratio of rNTP to dNTP in the reaction mixes was between 1:80 and 1:8. Primer extension reactions were performed in a buffer consisting of 50 mM Bicine (N,N-bis (2-hydroxyethyl)glycine; pH 8.3), 25 mM KOAc, and 2.5 mM MgCl₂. Four individual reactions, one for each of the four rNTPs, were performed. Each reaction (50 μ l) contained 200 μ M each dATP, dCTP, dGTP and dTTP (Perkin-Elmer) and 0.09 pmoles M13mp18 single-strand DNA template (Perkin-Elmer) annealed to 5'- 32 P] labeled DG48 (Lawyer et al., 1993, *PCR Methods and Applications* 2:275-287). The reactions also contained 2.5, 2.5, 2.5 or 25 μ M rATP, rCTP, rGTP or rUTP, respectively.

Each of the four reactions was initiated by addition of 7 units of G46D E615G Taq DNA polymerase and incubated at 75°C for 10 minutes. The reactions were stopped by addition of 10 μ l 60 mM EDTA and placed on ice. Twenty μ l of each reaction were added to 80 μ l of 50 mM Bicine (pH 8.3), 25 mM KOAc, and 2.5 mM MgCl₂. Cleavage products were produced by addition of 7 μ l of 1N NaOH and incubation for 15 minutes at 98°C. The reactions were neutralized by addition of 7 μ l of 1N HCl. Each reaction was precipitated by the addition of 312 μ l 95% ethanol and 10 μ l 3 M sodium acetate (pH 4.8). The reactions were microcentrifuged for 15 minutes to collect precipitate, the supernatant was removed, the pellets were washed with 500 μ l 70% ethanol and dried. Each pellet was resuspended in 5 μ l of 0.5% Stop Buffer (available from Perkin Elmer, Norwalk CT; contains 95% formamide, 20 mM EDTA and 0.05% bromophenol blue), heated at 98°C for 3 minutes, and directly loaded onto a pre-electrophoresed 6% polyacrylamide/8 M urea DNA sequencing gel and electrophoresed. The gel was dried and exposed to X-ray film. The resulting film revealed a clear sequencing ladder which provided in excess of 100 bases of correct sequence.

Example V

45 DNA Sequencing Using G46D, E615G, F667Y
Taq DNA Polymerase and 3'-deoxy Nucleotide Triphosphates

This example demonstrates the application of the modified polymerase, G46D, E615G, F667Y Taq to DNA sequencing using 3'deoxy nucleotide triphosphates. This experiment was performed using 3'deoxy ATP; however, it could be extended to use with the other 3'deoxy nucleotides as well. Primer extension reactions were performed in a buffer consisting of 50 mM Bicine (pH 8.3), 25 mM KOAc, and 2.5 mM MgCl₂. Each reaction (50 μ l) contained 200 μ M each dATP, dCTP, dGTP and dTTP (Perkin-Elmer) and 0.09 pmoles M13mp18 single-strand DNA template (Perkin-Elmer) annealed to 5'- 32 P] labeled DG48 (Lawyer et al., 1993, *PCR Methods and Applications* 2: 275-287). The reactions also contained 0, 0.1, 0.25, 0.5, 1, or 5 μ M 3'deoxy ATP.

EP 0 823 479 A2

Each of the reactions was initiated by addition of 7 units of G46D, E615G, F667Y Taq DNA polymerase and incubated at 75°C for 10 minutes. The reactions were stopped by addition of 10 µl 60 mM EDTA and placed on ice. Thirty µl of each reaction was ethanol precipitated and resuspended in Stop Buffer, heated at 98°C for 3 minutes, and directly loaded onto a pre-electrophoresed 6% polyacrylamide/8 M urea DNA sequencing gel and electrophoresed. The gel was dried and exposed to X-ray film. The lanes which contained reactions done in the presence of cordycepin contained clearly discernible termination ladders. The lanes containing the most cordycepin, i.e. 5 µM, showed a termination ladder in which, on average, the bands were shorter in length than the lanes in which the cordycepin levels were lower. The lane containing the reaction done in the absence of cordycepin, showed mostly full-length product and no termination ladder. These results indicate that the mutant enzyme is able to incorporate cordycepin and incorporation of this molecule into a primer extension product causes termination. This method could also be used to create a DNA sequencing ladder, with 3'deoxy CTP, 3'deoxy GTP and 3'deoxy UTP as well.

Example VI

15 Dye Primer PCR Sequencing with G46D E615G Tag DNA Polymerase

This example demonstrates the application of the modified polymerase of the invention to dye primer sequencing, utilizing ribonucleoside triphosphates (rNTPs) in PCR and a ratio of rNTP:dNTP of no more than 1:30. Four individual reactions, one for each of the rNTPs, were performed. PCR sequencing reactions were performed in a buffer consisting of 25 mM Tris-HCl (pH 9), 5.0 mM MgCl₂, and 10% glycerol (v/v). Each reaction also contained 500 µM each dATP, dCTP, dGTP, dTTP (Perkin Elmer), 5x10³ copy/µl pBSM13+ plasmid (Stratagene) template linearized with *Xba*I restriction endonuclease, and 0.05 unit/µl G46D E615G Tag DNA polymerase. Ribo-ATP reactions (10 µl) contained 2.5 µM ATP (Pharmacia Biotech), 0.1 µM JOE M13 Reverse Dye Primer (Perkin Elmer), and 0.1 µM primer ASC46 (5'-CGCCATTGCCATTCCAG). Ribo-CTP reactions (10 µl) contained 2.5 µM CTP (Pharmacia Biotech), 0.1 µM FAM M13 Reverse Dye Primer (Perkin Elmer), and 0.1 µM primer ASC46. Ribo-GTP reactions (10 µl) contained 2.5 µM GTP (Pharmacia Biotech), 0.1 µM TAMRA M13 Reverse Dye Primer (Perkin Elmer), and 0.1 µM primer ASC46. Ribo-UTP reactions (20 µl) contained 16 µM UTP (Pharmacia Biotech), 0.1 µM ROX M13 Reverse Dye Primer (Perkin Elmer), and 0.1 µM primer ASC46.

Each of the four reactions were placed in a preheated (75°C) Perkin Elmer GeneAmp® PCR System 9600 thermal cycler and subjected to 30 cycles of 95°C for 10 seconds, 75°C for 10 seconds, 1 minute ramp to 65°C, and 65°C for 5 minutes. The rATP and rCTP reactions each generated 6 x 10¹¹ copies of dye-labeled amplified 300 base pair product, and the rGTP and UTP reactions each generated 1.2 x 10¹² copies of dye-labeled amplified 300 base pair product.

To determine the DNA sequence of the amplified PCR products without requiring a separate enzymatic DNA sequencing reaction, the reactions were pooled, treated with base and heat, neutralized, and precipitated as follows. Four µl each of the ATP and CTP reactions and 8 µl each of the GTP and UTP reactions were pooled. Two microliters of 0.25 M EDTA (pH 8.0) (10 mM final), 10 µl 1 M NaOH (200 mM final), and 14 µl H₂O were added to the pooled reaction which was then incubated at 95°C for 5 minutes in a GeneAmp® PCR System 9600 thermal cycler and neutralized with 10 µl 1 M HCl. The pooled reaction was then precipitated by the addition of 150 µl 95% ethanol followed by an incubation at 4°C for 15 minutes. It was then microcentrifuged for 15 minutes at 4°C to collect the precipitate, and the supernatant removed by aspiration. The pellet was washed with 300 µl 70% ethanol, microcentrifuged for 5 minutes, the supernatant removed by aspiration, and the pellet dried. The pellet was resuspended in 6 µl formamide 50 mg/ml Blue dextran (in 25 mM EDTA) 5:1 (v/v) and heated at 90°C for 3 minutes. One and a half µl of the resuspended pellet was directly loaded onto a pre-electrophoresed 5% Long Ranger (FMC BioProducts), 6 M urea sequencing gel. It was then electrophoresed and analyzed on a Perkin Elmer ABI Prism™ 377 DNA Sequencer according to the manufacturers instructions. Automated base-calling by the Perkin Elmer ABI Prism™ Sequencing Analysis software resulted in greater than 99% accuracy for DNA sequence determination of the PCR amplified 300 base pair product.

SEQUENCE LISTING

5 (1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: F.Hoffmann-La Roche Ltd
- (B) STREET: Grenzacherstrasse 124
- (C) CITY: Basel
- (D) STATE: BS
- (E) COUNTRY: Switzerland
- (F) POSTAL CODE (ZIP): CH-4070
- (G) TELEPHONE: (0)61 688 24 03
- (H) TELEFAX: (0)61 688 13 95
- (I) TELEX: 962292/965512 hlr ch

(ii) TITLE OF INVENTION: Modified thermostable DNA polymerase

20 (iii) NUMBER OF SEQUENCES: 8

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 60/023,376
- (B) FILING DATE: 06-AUG-1996

30 (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40 (ix) FEATURE:

- (A) NAME/KEY: peptide
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Xaa
/note= "wherein Xaa is any amino acid but not Glu"

(ix) FEATURE:

- (A) NAME/KEY: peptide
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Xaa
/note= "wherein Xaa is Ile or Val"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Ser Gln Ile Xaa Leu Arg Xaa
1 5

6

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15

(ix) FEATURE:
(A) NAME/KEY: peptide
(B) LOCATION: 7
(D) OTHER INFORMATION: /label= Xaa
/note= "wherein Xaa is Ile or Val"

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ser Gln Ile Glu Leu Arg Xaa
1 5

25

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ser Gln Ile Glu Leu Arg Val
1 5

40

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Ser Gln Ile Glu Leu Arg Ile
1 5

55

EP 0 823 479 A2

(2) INFORMATION FOR SEQ ID NO:5:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
1 5 10 15

18 (2) INFORMATION FOR SEQ ID NO:6:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: peptide

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Leu Asp Tyr Ser Gln Ile Gly Leu Arg Val Leu Ala His Leu Ser
1 5 10 15

30 (2) INFORMATION FOR SEQ ID NO: 7:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2626 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

40 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Thermus aquaticus*

45 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 121..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

50 AAGCTCAGAT CTACCTGCCT GAGGGCGTCC GGTTCAGCT GGCCCTTCCC GAGGGGGAGA 60
GGGAGGCGT TCTAAAAGCC CTTCAGGACG CTACCCGGGG CGCGGTGGTG GAAGGGTAAC 120

ATG	AGG	GGG	ATG	CTG	CCC	CTC	TTT	GAG	CCC	AAG	GGC	CGG	GTC	CTC	CTG	168
Met	Arg	Gly	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	
1			5					10					15			
5																
GTG	GAC	GGC	CAC	CAC	CTG	GCC	TAC	CGC	ACC	TTC	CAC	GCC	CTG	AAG	GGC	216
Val	Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	His	Ala	Leu	Lys	Gly	
20								25					30			
10																
CTC	ACC	ACC	AGC	CGG	GGG	GAG	CGG	GTC	CAG	GCG	GTC	TAC	GGC	TTC	GCC	264
Leu	Thr	Thr	Sec	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala	
35							40					45				
15																
AAG	AGC	CTC	CTC	AAG	GCC	CTC	AAG	GAG	GAC	GGG	GAC	GCG	GTG	ATC	GTG	312
Lys	Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Asp	Ala	Val	Ile	Val	
50							55					60				
15																
GTC	TTT	GAC	GCC	AAG	GCC	CCC	TCC	TTC	CGC	CAC	GAG	GCC	TAC	GGG	GGG	360
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Glu	Ala	Tyr	Gly	Gly	
65							70					75		80		
20																
TAC	AAG	GCG	CGG	GCC	CCC	ACG	CCG	GAG	GAC	TTT	CCC	CGG	CAA	CTC	408	
Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu	
85							90					95				
25																
GCC	CTC	ATC	AAG	GAG	CTG	GTG	GAC	CTC	CTG	GGG	CTG	GCG	CGC	CTC	GAG	456
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp									
100							105					110				
25																
GTC	CCG	GCC	TAC	GAG	GCG	GAC	GTC	CTG	GCC	AGC	CTG	GCC	AAG	AAG	504	
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Ser	Leu	Ala	Lys	Lys	
115							120					125				
30																
GCG	GAA	AAG	GAG	GGC	TAC	GAG	GTC	CGC	ATC	CTC	ACC	GCC	GAC	AAA	GAC	552
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Lys	Asp	
130							135					140				
35																
CTT	TAC	CAG	CTC	CTT	TCC	GAC	CGC	ATC	CAC	GTC	CTC	CAC	CCC	GAG	GGG	600
Leu	Tyr	Gln	Leu	Leu	Ser	Asp	Arg	Ile	His	Val	Leu	His	Pro	Glu	Gly	
145							150					155		160		
35																
TAC	CTC	ATC	ACC	CGG	GCC	TGG	CTT	TGG	GAA	AAG	TAC	GGC	CTG	AGG	CCC	648
Tyr	Leu	Ile	Thr	Pro	Ala	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	
165							170					175				
40																
GAC	CAG	TGG	GCC	GAC	TAC	CGG	GCC	CTG	ACC	GGG	GAC	GAG	TCC	GAC	AAC	696
Asp	Gln	Trp	Ala	Asp	Tyr	Arg	Ala	Leu	Thr	Gly	Asp	Glu	Ser	Asp	Asn	
180							185					190				
45																
GTC	CCC	GGG	GTC	AAG	GGC	ATC	GGG	GAG	AAG	ACG	GCG	AGG	AAG	CTT	CTG	744
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Arg	Lys	Leu	Leu	
195							200					205				
45																
GAG	GAG	TGG	GGG	AGC	CTG	GAA	GCC	CTC	CTC	AAG	AAC	CTG	GAC	CGG	CTG	792
Glu	Glu	Trp	Gly	Ser	Leu	Glu	Ala	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Leu	
210							215					220				
50																
AAG	CCC	GCC	ATC	CGG	GAG	AAG	ATC	CTG	GCC	CAC	ATG	GAC	GAT	CTG	AAG	840
Lys	Pro	Ala	Ile	Arg	Glu	Lys	Ile	Leu	Ala	His	Met	Asp	Asp	Leu	Lys	
225							230					235		240		

EP 0 823 479 A2

5	CTC TCC TGG GAC CTG GCC AAG GTG CGC ACC GAC CTG CCC CTG GAG GTG Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu Val 245 250 255	888
10	GAC TTC GCC AAA AGG CGG GAG CCC GAC CGG GAG AGG CTT AGG GCC TTT Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala Phe 260 265 270	936
15	CTG GAG AGG CTT GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC CTT CTG Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu 275 280 285	984
20	GAA AGC CCC AAG GCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA GGG Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly 290 295 300	1032
25	GCC TTC GTG GGC TTT GTG CTT TCC CCC AAG GAG CCC ATG TGG GCC GAT Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala Asp 305 310 315 320	1080
30	CTT CTG GCC CTG GCC AGG GGG GGC CGG GTC CAC CGG GCC CCC Leu Leu Ala Leu Ala Ala Arg Gly Gly Arg Val His Arg Ala Pro 325 330 335	1128
35	GAG CCT TAT AAA GCC CTC AGG GAC CTG AAG GAG GCG CGG GGG CTT CTC Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu 340 345 350	1176
40	GCC AAA GAC CTG AGC GTT CTG GCC CTG AGG GAA GCC CTT GGC CTC CCG Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro 355 360 365	1224
45	CCC GGC GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCT TCC AAC Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn 370 375 380	1272
50	ACC ACC CCC GAG GGG GTG GCC CGG CGC TAC GGC GGG GAG TGG ACG GAG Thr Thr Pro Glu Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu 385 390 395 400	1320
55	GAG GCG GGG GAG CGG GCC CTT TCC GAG AGG CTC TTC GCC AAC CTG Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn Leu 405 410 415	1368
60	TGG GGG AGG CTT GAG GGG GAG GAG AGG CTC CTT TGG CTT TAC CGG GAG Trp Gly Arg Leu Glu Gly Glu Arg Leu Leu Trp Leu Tyr Arg Glu 420 425 430	1416
65	CTG GAG AGG CCC CTT TCC GCT GTC CTG GCC CAC ATG GAG GCC ACG GGG Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly 435 440 445	1464
70	CTG CGC CTG GAC GTG GCC TAT CTC AGG GCC TTG TCC CTG GAG GTG GCC Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala 450 455 460	1512

EP 0 823 479 A2

	GAG GAG ATC GCC CGC CTC GAG GCC GAG GTC TTC CGC CTG GCC GGC CAC	1560
	Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His	
465	470	475
	480	
5	CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTC CTC TTT GAC	1608
	Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp	
	485	490
	495	
10	GAG CTA GGG CTT CCC GCC ATC GGC AAG ACC GAG AAG ACC GGC AAG CGC	1656
	Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg	
	500	505
	510	
15	TCC ACC AGC GCC GCC GTC CTG GAG GCC CTC CGC GAG GCC CAC CCC ATC	1704
	Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile	
	515	520
	525	
20	GTG GAG AAG ATC CTG CAG TAC CGG GAG CTC ACC AAG CTG AAG AGC ACC	1752
	Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr	
	530	535
	540	
25	TAC ATT GAC CCC TTG CCG GAC CTC ATC CAC CCC AGG ACG GGC CGC CTC	1800
	Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu	
	545	550
	555	560
30	CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG GGC AGG CTA AGT AGC	1848
	His Thr Arg Phe Asn Gln Thr Ala Thr Gly Arg Leu Ser Ser	
	565	570
	575	
35	TCC GAT CCC AAC CTC CAG AAC ATC CCC GTC CGC ACC CCG CTT GGG CAG	1896
	Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln	
	580	585
	590	
40	AGG ATC CGC CGG GCC TTC ATC GCC GAG GAG GGG TGG CTA TTG GTG GCC	1944
	Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala	
	595	600
	605	
45	CTG GAC TAT AGC CAG ATA GGG CTC AGG GTG CTG GCC CAC CTC TCC GGC	1992
	Leu Asp Tyr Ser Gln Ile Gly Leu Arg Val Leu Ala His Leu Ser Gly	
	610	615
	620	
50	GAC GAG AAC CTG ATC CGG GTC TTC CAG GAG GGG CGG GAC ATC CAC ACG	2040
	Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr	
	625	630
	635	640
55	GAG ACC GCC AGC TGG ATG TTC GGC GTC CCC CGG GAG GCC GTG GAC CCC	2088
	Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro	
	645	650
	655	
60	CTG ATG CGC CGG GCG GCC AAG ACC ATC AAC TTC GGG GTC CTC TAC GGC	2136
	Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly	
	660	665
	670	
65	ATG TCG GCC CAC CGC CTC TCC CAG GAG CTA GCC ATC CCT TAC GAG GAG	2184
	Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu	
	675	680
	685	
70	GCC CAG GCC TTC ATT GAG CGC TAC TTT CAG AGC TTC CCC AAG GTG CGG	2232
	Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg	
	690	695
	700	

EP 0 823 479 A2

5	GCC TGG ATT GAG AAG ACC CTG GAG GAG GGC AGG AGG CGG GGG TAC GTG Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val 705 710 715 720	2280
10	GAG ACC CTC TTC GGC CGC CGC CGC TAC GTG CCA GAC CTA GAG GCC CGG Glu Thr Leu Phe Gly Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg 725 730 735	2328
15	GTC AAG AGC GTG CGG GAG GCG GCC GAG CGC ATG GCC TTC AAC ATG CCC Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro 740 745 750	2376
20	GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTG GCT ATG GTG AAG CTC Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu 755 760 765	2424
25	TTC CCC AGG CTG GAG GAA ATG GGG GCC AGG ATG CTC CTT CAG GTC CAC Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His 770 775 780	2472
30	GAC GAG CTG GTC CTC GAG GCC CCA AAA GAG AGG GCG GAG GCC GTG GCC Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala 785 790 795 800	2520
35	CGG CTG GCC AAG GAG GTC ATG GAG GGG GTG TAT CCC CTG GCC GTG CCC Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro 805 810 815	2568
40	CTG GAG GTG GAG GTG GGG ATA GGG GAG GAC TGG CTC TCC GCC AAG GAG Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu 820 825 830	2616
45	TGATACCACCC	2626

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
35 (A) LENGTH: 832 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
1 5 10 15
45 Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly
20 25 30
Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
35 40 45
50 Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile Val
55 60

EP 0 823 479 A2

Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly Gly
 65 70 75 80
 5 Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
 85 90 95
 Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu Glu
 100 105 110
 10 Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys Lys
 115 120 125
 Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys Asp
 130 135 140
 15 Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu Gly
 145 150 155 160
 Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro
 165 170 175
 20 Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp Asn
 180 185 190
 Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu Leu
 195 200 205
 25 Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg Leu
 210 215 220
 Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu Lys
 225 230 235 240
 30 Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu Val
 245 250 255
 Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala Phe
 260 265 270
 35 Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu
 275 280 285
 Glu Ser Pro Lys Ala Leu Glu Ala Pro Trp Pro Pro Pro Glu Gly
 290 295 300 300
 40 Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala Asp
 305 310 315 320
 Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala Pro
 45 325 330 335
 Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu
 340 345 350
 Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro
 50 355 360 365
 Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn
 370 375 380

EP 0 823 479 A2

Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu
 385 390 395 400
 5 Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn Leu
 405 410 415
 Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg Glu
 420 425 430
 10 Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly
 435 440 445
 Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala
 450 455 460
 15 Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His
 465 470 475 480
 Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp
 485 490 495
 20 Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg
 500 505 510
 Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile
 515 520 525
 25 Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr
 530 535 540
 Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu
 545 550 555 560
 30 His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser
 565 570 575
 Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln
 580 585 590
 35 Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala
 595 600 605
 Leu Asp Tyr Ser Gln Ile Gly Leu Arg Val Leu Ala His Leu Ser Gly
 610 615 620
 40 Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr
 625 630 635 640
 Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro
 45 645 650 655
 Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly
 660 665 670
 50 Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu
 675 680 685
 Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg
 690 695 700

Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val
 705 710 715 720

5 Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg
 725 730 735

Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro
 740 745 750

10 Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu
 755 760 765

Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His
 770 775 780

15 Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala
 785 790 795 800

Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro
 805 810 815

20 Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
 820 825 830

25

Claims

1. A thermostable DNA polymerase enzyme that comprises the amino acid sequence SerGlnIleXaaLeuArgXaa (SEQ ID NO: 1), wherein "Xaa" at position 4 of this sequence is any amino acid residue but not a glutamic acid residue (Glu) and is preferably a glycine (Gly) residue, and wherein "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile).
2. The thermostable DNA polymerase enzyme of claim 1 that is characterized in that it is a recombinant derivative of a naturally occurring thermostable DNA polymerase, wherein said naturally occurring thermostable DNA polymerase comprises the amino acid sequence motif SerGlnIleGluLeuArgXaa (SEQ ID NO: 2), wherein "Xaa" at position 7 of this sequence is a valine residue (Val) or an isoleucine residue (Ile).
3. The thermostable DNA polymerase enzyme of claim 2 that is characterized in that it displays reduced discrimination against incorporation of an unconventional nucleotide in comparison to said naturally occurring thermostable DNA polymerase.
4. The thermostable DNA polymerase of claim 3 further characterized in that the ability of said polymerase to incorporate an unconventional nucleotide, relative to the ability of said corresponding native form of polymerase to incorporate said unconventional nucleotide, is increased by at least 20 fold.
5. The thermostable DNA polymerase as claimed in any one of claims 1 to 4 characterized in that said polymerase has sufficient activity for use in a DNA sequencing reaction that comprises an unconventional nucleotide which is preferably a ribonucleoside triphosphate and a corresponding conventional nucleotide in a ratio of 1:1 or less.
6. The thermostable DNA polymerase as claimed in any one of claims 1 to 4 characterized in that said polymerase has sufficient activity for use in a DNA sequencing reaction that comprises an unconventional nucleotide which is a ribonucleoside triphosphate present at a concentration of less than about 100 μ M and a corresponding conventional nucleotide which is present at a concentration of more than about 100 μ M.
7. The thermostable DNA polymerase enzyme of any one of claims 2 to 6 which is a recombinant derivative of a naturally occurring thermostable DNA polymerase enzyme from an organism selected from the group consisting of *Thermus aquaticus*, *Thermus caldophilus*, *Thermus chlorophilus*, *Thermus filiformis*, *Thermus flavus*, *Thermus*

EP 0 823 479 A2

oshimai, Thermus ruber, Thermus scotoductus, Thermus silvanus, Thermus species Z05, Thermus species sps17, Thermus thermophilus, Thermotoga maritima, Thermotoga neopolitana, Thermosiphon africanus, Anaerocellum thermophilum, Bacillus caldoterax, and Bacillus stearothermophilus.

5 8. The thermostable DNA polymerase enzyme of any one of claims 2 to 6 which is a recombinant derivative of a naturally occurring thermostable Thermus species DNA polymerase, preferably of Taq DNA polymerase or a homologous polymerase thereof, more preferably of a thermostable DNA polymerase comprising the amino acid sequence LeuAspTyrSerGlnIleGlu.euArgValLeuAlaHisLeuSer (SEQ ID NO: 5), most preferably the thermostable DNA polymerase having the amino acid sequence of SEQ ID NO: 8 and modified forms thereof, such as the G46D and/or the F66Y mutant forms.

10 9. The thermostable DNA polymerase enzyme of any one of claims 1 to 6 which has at least about 39 %, preferably at least about 60 %, more preferably at least about 80 % sequence homology to the amino acid sequence of Taq DNA polymerase (SEQ ID NO: 7).

15 10. A nucleic acid sequence encoding a thermostable DNA polymerase enzyme as claimed in any one of claims 1 to 9.

11. A vector comprising a nucleic acid sequence encoding a thermostable DNA polymerase enzyme as claimed in any one of claims 1 to 9.

20 12. A host cell comprising a nucleic acid sequence encoding a thermostable DNA polymerase enzyme as claimed in any one of claims 1 to 9.

13. A method for preparing a thermostable DNA polymerase enzyme, comprising:

25 (a) culturing a host cell of claim 12 under conditions which promote the expression of the thermostable DNA polymerase enzyme; and
(b) isolating the thermostable DNA polymerase enzyme from the host cell or from the culture medium.

30 14. A thermostable DNA polymerase enzyme prepared by the method as claimed in claim 13.

15. Use of a thermostable DNA polymerase enzyme as claimed in any one of claims 1 to 9 in a nucleic acid amplification or sequencing reaction.

35 16. A composition for use in a DNA sequencing reaction that comprises; a nucleic acid template; an oligonucleotide primer complementary to said template; a thermostable DNA polymerase as claimed in any one of claims 1 to 9, a mixture of conventional dNTPs, and at least one unconventional nucleotide, wherein the ratio of said unconventional nucleotide to said corresponding conventional nucleotide is 1:1 or less.

40 17. The composition of claim 16 wherein said unconventional nucleotide is a rhonucleotide, whereby said rhonucleotide is preferably present at a concentration of less than about 100 μ M and the corresponding conventional nucleotide is present at a concentration of more than about 100 μ M.

45 18. The composition of claim 17 further characterized in that said unconventional nucleotide is unlabeled.

19. A method for sequencing a nucleic acid target which method comprises the steps of:

50 (a) providing an unconventional nucleotide and a corresponding conventional nucleotide in a DNA sequencing reaction, wherein said unconventional and corresponding conventional nucleotides are present in a ratio of less than about 1:1;
(b) treating the reaction of step (a) in the presence of a thermostable DNA polymerase as claimed in any one of claims 1 to 9 under conditions for primer extension to provide primer extension products comprising said unconventional nucleotide;
(c) treating the primer extension products of step (b) under conditions for hydrolyzing said primer extension products;
(d) resolving reaction products from step (c); and
(e) determining the sequence of the nucleic acid target.

EP 0 823 479 A2

20. The method for sequencing of claim 19 wherein said unconventional nucleotide is a ribonucleotide, which is preferably present at a concentration of about 0.1 μ M - 100 μ M.
- 5 21. The method for sequencing of claim 19 wherein said corresponding conventional nucleotide is present at a concentration of about 50 μ M - 500 μ M.
- 10 22. A kit for sequencing a nucleic acid comprising a thermostable DNA polymerase as claimed in any one of claims 1 to 9 and optionally further reagents useful in such sequencing procedure such as e.g. one or more oligonucleotide primers, a mixture of conventional dNTPs, and at least one unconventional nucleotide, wherein the ratio of said unconventional nucleotide to said corresponding conventional nucleotide is preferably less than one.

15

20

25

30

35

40

45

50

55